10-3-2016

Matching the Efficiency Gains of the Logistic Regression Estimator While Avoiding its Interpretability Problems, in Randomized Trials

Michael Rosenblum
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, mrosenbl@jhsph.edu

Jon Arni Steingrimsson
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics

Matching the Efficiency Gains of the Logistic Regression Estimator while Avoiding its Interpretability Problems, in Randomized Trials

MICHAEL ROSENBLUM
Department of Biostatistics,
Johns Hopkins Bloomberg School of Public Health,
Baltimore, Maryland, 21205, USA.
mrosen@jhu.edu

JON ARNI STEINGRIMSSON
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
Baltimore, Maryland, 21205, USA.
jsteing5@jhu.edu

Abstract

Adjusting for prognostic baseline variables can lead to improved power in randomized trials. For binary outcomes, a logistic regression estimator is commonly used for such adjustment. This has resulted in substantial efficiency gains in practice, e.g., gains equivalent to reducing the required sample size by 20–28% were observed in a recent survey of traumatic brain injury trials. Robinson and Jewell (1991) proved that the logistic regression estimator is guaranteed to have equal or better asymptotic efficiency compared to the unadjusted estimator (which ignores baseline variables). Unfortunately, the logistic regression estimator has the following dangerous vulnerabilities: it is only interpretable when the treatment effect is identical within every stratum of baseline covariates; also, it is inconsistent under model misspecification, which is virtually guaranteed when the baseline covariates are continuous or categorical with many levels. An open problem was whether there exists an equally powerful, covariate-adjusted estimator with no such vulnerabilities, i.e., one that (i) is interpretable and consistent without requiring any model assumptions, and (ii) matches the efficiency gains of the logistic regression estimator. Such an estimator would provide the best of both worlds: interpretability and consistency under no model assumptions (like the unadjusted estimator) and power gains from covariate adjustment (that match the logistic regression estimator). We prove a new asymptotic result showing that, surprisingly, there are simple estimators satisfying the above properties. We argue that these rarely used estimators have substantial advantages over the more commonly used logistic regression estimator for covariate adjustment in randomized trials with binary outcomes. Though our focus is binary outcomes and logistic regression models, our results extend to a large class of generalized linear models.

Keywords: Pitman Efficiency; Robustness; Hypothesis Test

1 Introduction

Austin et al. (2010) conducted a review of randomized clinical trial reports from medical journals, and concluded that "There is a need for an informed debate about the relative interpretability and utility for clinical and policy decision making of unadjusted vs. adjusted measures of treatment effect for binary and time-to-event outcomes." We aim to contribute to this debate, focusing on binary outcomes. Our goal is to recommend an estimator (with corresponding confidence interval and hypothesis test) for the primary efficacy analysis of a confirmatory randomized trial.

Adjusting for prognostic baseline variables can lead to efficiency gains, as illustrated by Hernández et al. (2006). They estimated the gains from adjusting for baseline covariates in seven phase 3 randomized trials of treatments for traumatic brain injury. The primary outcome in each trial was the dichotomized Glasgow Outcome Scale of functional disability measured at 6 months. Prognostic baseline variables included "age, motor score, pupillary reactivity, computed tomography (CT) classification, traumatic subarachnoid hemorrhage, hypoxia, hypotension, glycemia, and hemoglobin" (Hernández et al., 2006). A logistic regression estimator was compared to the unadjusted estimator. The former, referred to below as the logistic coefficient estimator, is the estimated coefficient on the treatment term in a main effects logistic regresssion model for the outcome given treatment and baseline variables. The unadjusted estimator is the difference between the sample proportions of successful outcomes in the two study arms. Efficiency gains from using the logistic coefficient estimator compared to the unadjusted estimator were equivalent to sample size reductions ranging from 20–28%.

We consider trials where each participant is randomized to the treatment or control arm, independent of baseline variables; extensions to stratified randomization are discussed in Section 6. Our focus is the intention-to-treat analysis, which compares the impact of assignment to the treatment versus control arm. The goal of the analysis is to estimate the average treatment effect, construct a confidence interval for it, and test the null hypothesis of no average treatment effect. We compare three estimators: the unadjusted, the logistic coefficient, and the standardized estimator. The latter two are adjusted estimators that leverage information in baseline variables, and are defined in Section 2.3. Each estimator can be used for hypothesis testing by dividing by its standard error and comparing to the appropriate quantile of the standard normal distribution.

Asymptotic relative efficiency (also called Pitman efficiency) is used to compare test statistics from different estimators. It represents the ratio of sample sizes required to achieve a desired power and Type I error, comparing two testing methods. We refer to asymptotic relative efficiency simply as "efficiency" for conciseness.

Robinson and Jewell (1991, Section 8) compared the efficiency of the logistic coefficient estimator versus the unadjusted estimator. They proved that the former has equal or better efficiency compared to the latter, assuming the logistic regression model is correctly specified.

It was an open problem to determine whether the logistic coefficient estimator or the standardized estimator is more efficient for testing the null hypothesis of no average treatment effect. This was unknown both for the case where the logistic regression model is correctly specified, and also for the case where the model is misspecified. The latter case may be most important, since in practice one would expect the model to be at least somewhat misspecified.

Our main contribution is proving that the standardized estimator has equal efficiency compared to the logistic coefficient estimator; this holds not only when the logistic regression model is correctly specified, but also under arbitrary model misspecification. Therefore, there is no advantage in terms of power gains, asymptotically, to using the logistic coefficient estimator compared to the standardized estimator. This is important since the latter estimator has substantial advantages compared to the former.

The main advantage of the standardized estimator is its interpretability. The population parameter estimated by the standardized estimator is the same as that estimated by the unadjusted estimator, i.e., the average treatment effect. This effect, also called a marginal or unconditional effect, has the direct interpretation as a contrast between the probability of a successful outcome if everyone in the target population were assigned to treatment versus control. If the average effect is positive, then giving the treatment to everyone in the target population would lead to better outcomes, on average, compared to control. The standardized estimator is guaranteed to converge to the average treatment effect, regardless of whether the logistic regression model is correctly specified or not.

The population parameter estimated by the logistic coefficient estimator is the conditional treatment effect within strata of the baseline variables. This estimator is only interpretable under the assumption that the conditional effect is identical within every such stratum, i.e., under the assumption that the conditional effect is a single number rather than a function that can vary depending on baseline variables (Freedman, 2008). Even if this assumption were true, the logistic coefficient estimator is inconsistent under misspecification of the logistic regression model.

In brief, our main result shows that the standardized estimator gets all the asymptotic efficiency gains of the logistic coefficient estimator without the interpretability and inconsistency problems of the latter. We focus on the standardized estimator of Moore and van der Laan (2009) due to its simplicity and ease of implementation. Our result also has implications for a variety of standardized estimators, e.g., those based on (Robins et al., 2007; Moore and van der Laan, 2009; Tan, 2010; Rotnitzky et al., 2012; Gruber and van der Laan, 2012; Colantuoni and Rosenblum, 2015) when applied to randomized trials, as discussed in Section 6.

2 Problem Definition

2.1 Data Structure and Assumptions

Let Y denote the binary outcome, A denote assignment to treatment ($A = 1$) or control ($A = 0$), and B denote a column vector of baseline variables which can be any mix of categorical and continuous variables measured before randomization. The baseline variables B must be prespecified in the study protocol. Each participant i has data vector (B_i, A_i, Y_i). A total of n participants are enrolled.

Each participant's data vector (B_i, A_i, Y_i) is assumed to be an independent, identically distributed draw from the unknown joint distribution P_0 on (B, A, Y). No assumptions are made on P_0 except that A is a Bernoulli draw with probability $1/2$ of being 0 or 1 independent of B (which holds by randomization), B is bounded, and P_0 satisfies regularity conditions given later in the paper.

2.2 Treatment Effect and Null Hypothesis Definitions

Define the unconditional probability of success under treatment and control to be $\mu_1 = P_0(Y = 1|A = 1)$ and $\mu_0 = P_0(Y = 1|A = 0)$, respectively. These probabilities are nonparametrically defined, i.e., they do not require any model assumptions (such as a logistic regression model) in order to be well-defined and interpretable. We focus throughout on testing the null hypothesis of no average treatment effect: $H_0 : \mu_1 = \mu_0$. The average treatment effect on the risk difference scale is defined as $\mu_1 - \mu_0$. The analogous average treatment effects on the relative risk and log-odds scales are defined as μ_1/μ_0 and $\text{logit}(\mu_1) - \text{logit}(\mu_0)$, respectively, for $\text{logit}(x) = \log\{x/(1-x)\}$.

Define the following logistic regression model for the outcome given study arm assignment and baseline variables:

$$\text{logit}\{P(Y = 1|A, B)\} = \gamma_0 + \gamma_1 A + \gamma_2' B, \tag{1}$$

where γ_2 is a column vector of same length as B, and γ_2' denotes its transpose. We do not assume this model is correctly specified. That is, the true joint distribution $P_0(B, A, Y)$ need not satisfy any of the restrictions encoded in this model (such as equal conditional treatment effect within every stratum of baseline variables, and the relationship between the outcome and baseline variables having the simple, linear form above for each study arm). Our only assumptions about P_0 are those in Section 2.1.

If the model (1) is correct, then H_0 is equivalent to the conditional null hypothesis $H_{0C}: P(Y=1|A=1,B=b) = P(Y=1|A=0,B=b)$ for every stratum b of baseline variables with positive density under P; this follows since if the model is correct, each null hypothesis is equivalent to $\gamma_1 = 0$. If the model is misspecified, then H_0 does not necessarily imply the sharper null hypothesis H_{0C}. In general, neither null hypothesis implies the model (1) is correctly specified. (An exception is when B is a single binary variable, in which case H_{0C} implies the model is correctly specified.) We focus on testing H_0 rather than H_{0C}, since the former is typically of primary interest in confirmatory randomized trials.

2.3 Estimators

The unadjusted estimator $\hat{\psi}_{unadj}$ of the marginal risk difference $\mu_1 - \mu_0$ is the difference between the sample proportions with $Y = 1$ between the treatment and control arms.

Let $\hat{\gamma} = (\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2)$ denote the estimated coefficients when (1) is fit using maximum likelihood estimation. The logistic coefficient estimator is defined as $\hat{\gamma}_1$. Even when the model (1) is misspecified, $\hat{\gamma}$ converges in probability to the maximizer $\bar{\gamma}$ of the expected log-likelihood

$$E_{P_0} \log \left\{\operatorname{expit}\left(\gamma_0 + \gamma_1 A + \gamma_2' B\right)\right\}^Y \left\{1 - \operatorname{expit}\left(\gamma_0 + \gamma_1 A + \gamma_2' B\right)\right\}^{1-Y},$$

where $\operatorname{expit} = \operatorname{logit}^{-1}$, E_{P_0}denotes expectation with respect to P_0, and we assume throughout that the expected log-likelihood has a unique maximizer. (Rosenblum and van der Laan (2009) showed that if the components of B are linearly independent, then the expected log-likelihood is strictly concave and so any local maximum is the unique, global maximum.) Under misspecification of (1), the probability limit $\bar{\gamma}_1$ of the logistic coefficeint estimator $\hat{\gamma}_1$ is generally uninterpretable.

The standardized estimator of the marginal risk difference is defined as:

$$\hat{\psi}_{std} = \frac{1}{n}\sum_{i=1}^{n} \operatorname{expit}(\hat{\gamma}_0 + \hat{\gamma}_1 + \hat{\gamma}_2' B_i) - \frac{1}{n}\sum_{i=1}^{n} \operatorname{expit}(\hat{\gamma}_0 + \hat{\gamma}_2' B_i). \tag{2}$$

This estimator is from Moore and van der Laan (2009), and is a special case of a class of estimators from Scharfstein et al. (1999). We emphasize that each of the sums in (2) is over all participants $i = 1, \ldots, n$ in the trial (not only those assigned to a specific arm, for example). The estimator can be thought of as standardizing to the marginal distribution of the baseline variables from the entire (pooled) data set.

The Wald statistic corresponding to each estimator is the estimator divided by its standard error. Throughout, we assume that a robust variance estimator (such as the nonparametric bootstrap) is used to compute each standard error, and that this robust variance estimator is consistent.

2.4 Validity of Hypothesis Tests Based on Each Estimator

It follows from Scharfstein et al. (1999) and Moore and van der Laan (2009) that under arbitrary misspecification of the model (1), both $\hat{\psi}_{unadj}$ and $\hat{\psi}_{std}$ converge in probability to $\mu_1 - \mu_0$ (the average treatment effect on the risk difference scale) and are asymptotically normal. Therefore, under H_0, these estimators converge to 0 in probability and their corresponding Wald statistics lead to tests of H_0 that have asymptotically correct Type I error; this holds regardless of whether the model (1) is correctly specified.

In contrast, the logistic coefficient estimator $\hat{\gamma}_1$ converges to the conditional effect $\text{logit}\{P(Y = 1|A = 1, B)\} - \text{logit}\{P(Y = 1|A = 0, B)\}$ (also called the conditional log-odds ratio) when (1) is correctly specified, and converges to an uninterpretable limit $\bar{\gamma}$ otherwise (Freedman, 2008). Gail et al. (1984) showed that when (1) is correctly specified and H_0 is false, the conditional effect has greater magnitude than the marginal effect on the log-odds scale. The logistic coefficient estimator $\hat{\gamma}_1$ is asymptotically normal, regardless of whether (1) holds.

Rosenblum and van der Laan (2009) proved that for testing H_{0C}, the Wald test based on the logistic coefficient estimator $\hat{\gamma}_1$ has asymptotically correct Type I error; this holds under arbitrary misspecification of the model (1). We strengthen this result, showing that the Wald test based on $\hat{\gamma}_1$ is also valid for testing the weaker null hypothesis H_0. This result and several results above are encapsulated in the following theorem:

Theorem 2.1. *For each estimator $\hat{\psi}_{unadj}$, $\hat{\gamma}_1$, $\hat{\psi}_{std}$, it converges to 0 in probability if and only if H_0 is true. Therefore, the Wald test of H_0 based on any of these estimators has asymptotically correct Type I error. These results hold under arbitrary misspecification of the model (1).*

All of our results are proved in the Appendix. We next compare the power of Wald tests based on the above estimators.

3 Main Result

Asymptotic relative efficiency, defined by Pitman (van der Vaart, 1998, p. 201), compares the large sample performance of two testing procedures. It represents the ratio of required sample sizes for each testing procedure to achieve a desired power and Type I error. The formal definition requires a set of alternatives $P^{(\nu)}(B, A, Y)$ (i.e., joint distributions) indexed by $\{\nu \in \mathbb{R} : \nu > 0\}$ that converge to some $P^{(0)} \in H_0$ as $\nu \downarrow 0$. The asymptotic relative efficiency for testing H_0 is the limit of the ratio of the minimum sample sizes needed by each testing procedure to achieve a desired power $1-\beta$ with Type I error at most α, under $P^{(\nu)}$ as $\nu \downarrow 0$. Our results hold for many possible sets of alternatives $\{P^{(\nu)} : \nu > 0\}$, and for any α, β satisfying $0 < \alpha < 1-\beta < 1$, under regularity conditions in the next paragraph.

We assume that $P^{(\nu)}$ converges in total variation distance to some $P^{(0)} \in H_0$ as $\nu \downarrow 0$, and that each $P^{(\nu)}$ satisfies the assumptions in Section 2.1. Define $\mu_a(\nu) = P^{(\nu)}(Y=1|A=a)$ for each $a \in \{0,1\}, \nu \geq 0$. For any $\nu > 0$, $P^{(\nu)}$ is assumed to satisfy the alternative hypothesis $\mu_1(\nu) - \mu_0(\nu) > 0$. (We focus on one-sided alternatives, but analogous results hold for the two-sided case.) Furthermore, we assume $\mu_1(\nu) - \mu_0(\nu)$ is right differentiable at $\nu = 0$, with positive right-derivative. Intuitively, this condition means that the parameter $\mu_1(\nu) - \mu_0(\nu)$, which defines the null hypothesis $H_0 : \mu_1(\nu) - \mu_0(\nu) = 0$, is increasing in ν (to first order) in small neighborhoods of $\nu = 0$. We assume the regularity conditions in Lemma 7.6 and Theorem 14.19 of van der Vaart (1998, pp. 95, 201). The former conditions imply that the parametrization $\{P^{(\nu)} : \nu \geq 0\}$ is regular, i.e., differentiable in quadratic mean. The latter conditions imply that asymptotic relative efficiency is determined by the slope of each Wald statistic, i.e., the ratio of the derivative of its asymptotic mean to its asymptotic dispersion.

An example of a set of alternatives satisfying the above assumptions is to let $P^{(\nu)}$ denote the distribution satisfying (1) at $\gamma_1 = \nu$ and with γ_0, γ_2, and the marginal distribution of B fixed (not changing with ν).

The following is our main result:

Theorem 3.1. *Consider any set of alternatives $\{P^{(\nu)} : \nu > 0\}$ satisfying the above regularity conditions. The asymptotic relative efficiency for testing H_0, comparing Wald statistics based on the standardized estimator versus the logistic coefficient estimator, is* 1.

The theorem shows that the standardized estimator is asymptotically as efficient as the logistic coefficient estimator. This holds regardless of whether the logistic regression model (1) is correctly specified; that is, we do not require $P^{(\nu)}$ or $P^{(0)}$ to satisfy (1).

Robinson and Jewell (1991) showed that the asymptotic relative efficiency of the logistic coefficient estimator compared to the unadjusted estimator is greater or equal to 1, assuming the model (1) is correct. In the Appendix, we slightly extend their result by showing it holds in the setting of our paper. Combining this with Theorem 3.1, we have:

Corollary 3.2. *Consider any $P^{(0)} \in H_0$ for which (1) is correctly specified, and any set of alternatives $\{P^{(\nu)} : \nu > 0\}$ that satisfy the above regularity conditions and converge to $P^{(0)}$ in total variation distance as $\nu \downarrow 0$. The asymptotic relative efficiency for testing H_0, comparing Wald statistics based on the standardized estimator versus the unadjusted estimator, is at least* 1.

(Asymptotic relative efficiency greater than 1 means that the first procedure requires smaller sample size compared to the second procedure, asymptotically.)

Theorems 2.1–3.1 and Corollary 1 involve the standardized estimator for the risk difference $\mu_1 - \mu_0$. These results still hold if the risk difference in the standardized estimator is replaced by any smooth contrast between μ_1 and μ_0, such as the relative risk reduction $1 - \mu_1/\mu_0$ or log-odds ratio $\text{logit}(\mu_1) - \text{logit}(\mu_0)$. In general, we define a smooth contrast between the marginal means μ_0, μ_1 to be any real-valued function r of (μ_0, μ_1) that is continuously differentiable, equals 0 under H_0 (i.e., whenever $\mu_0 = \mu_1$), and has gradient with non-zero magnitude. The corresponding standardized estimator involves substituting the first and second terms on the right side of (2) for μ_1 and μ_0, respectively, in the contrast r.

Theorem 3.3. *The standardized estimator using any smooth contrast r converges to 0 in probability if and only if H_0 is true. The asymptotic relative efficiency for testing H_0, comparing Wald statistics based on the standardized estimator using the risk difference versus the standardized estimator using any smooth contrast r, is 1.*

4 Simulation Study Based on the MISTIE II trial

The MISTIE II trial (Hanley et al., In Press) is a randomized phase II trial comparing a surgical procedure that removes blood clots to the standard of care in patients who have intracerebral hemorrhage. The primary outcome has value 1 if the participant's modified Rankin scale score at 180 days is 3 or less, and is 0 otherwise. We use the following prognostic baseline variables: age, intracerebral hemorrhage volume, and National Institutes of Health Stroke Scale. We use data from the 89 participants (out of 96 total) in the trial who have all of these variables and the outcome measured. The unadjusted estimate of the average treatment effect on the risk difference scale is 0.12; it is 0.54 on the log-odds scale. Our simulation study setup is similar to (Colantuoni and Rosenblum, 2015); however, they did not consider the logistic coefficient estimator, whose relative efficiency compared to the standardized estimator is the focus of our paper.

We compare the efficiency of the unadjusted, logistic coefficient, and standardized estimators by simulating 10000 trials, each with sample size 89. The data generating mechanism is constructed to mimic the correlation structure between the outcome and baseline variables in the MISTIE II data. Each simulated trial data set involves first sampling 89 pairs (Y, B) with replacement from the MISTIE II data. Each simulated participant's treatment assignment A is set to be treatment or control with probability 0.5 independent of (Y, B). In order to induce a positive average treatment effect equal to that observed in the MISTIE II trial, we modify some of the simulated participants' outcomes. Specifically, for each simulated participant with

$A = 1$ and $Y = 0$, we change Y to 1 with probability 0.17 based on a random draw independent of B. The resulting data generating distribution has average treatment effect 0.12 on the risk difference scale, and average treatment effect 0.54 on the log-odds scale. Under this distribution, the logistic regression model (1) is misspecified.

The relative efficiency comparing Wald statistics based on two estimators is approximated below by the ratio of the first estimator's signal to noise ratio (defined as the square of its mean divided by its variance) compared to that of the second estimator. Each mean and variance is approximated by its empirical mean and variance over the 10000 simulated trials.

Table 1 shows results from the 10000 simulated trials. The unadjusted and standardized estimators are calculated both on the risk difference and log-odds scales. For each estimator, the table shows its empirical mean and standard error. Relative efficiency compares the Wald test of H_0 based on each estimator to the Wald test of H_0 based on the logistic coefficient estimator.

	Risk Difference Scale		Log Odds Scale		
	Unadjusted	Standardized	Unadjusted	Standardized	Logistic
Mean	0.12	0.12	0.54	0.54	0.70
Standard Error	0.10	0.09	0.46	0.41	0.54
Relative Efficiency	0.86	1.04	0.83	1.02	1.00

Table 1: Comparison of the unadjusted, standardized, and logistic coefficient estimators.

The results in Table 1 show that both the unadjusted and standardized estimator are approximately unbiased for the average treatment effect. The relative efficiency results in Table 1 agree well with the results from Theorem 3.1 and Corollary 1. The unadjusted estimator is less efficient than both the standardized and the logistic coefficient estimators. The relative efficiency of the standardized versus the logistic coefficient estimator is close to 1. We also conducted simulations as above except with sample size 500; the relative efficiency of the standardized versus the logistic coefficient estimator becomes closer to one, with relative efficiencies of 1.01 and 0.99 when the standardized estimator is on the risk difference and log-odds scale, respectively. Consistent with Theorem 3.3, the relative efficiencies of the standardized estimator in Table 1 are similar for both the risk difference and log-odds scales.

5 Generalized Linear Models with Canonical Link Functions

Our results hold not only for binary outcomes and logistic regression models, but also for generalized linear models (GLM) with canonical link function (denoted g), under regularity conditions given below. This includes the following special cases from McCullagh and Nelder (1989):

- Linear regression for continuous outcomes ($g(x) = x$).
- Logistic regression for binary outcomes ($g(x) = \text{logit}(x)$).
- Poisson regression for count outcomes ($g(x) = \log(x)$).
- Gamma regression for positive, real-valued outcomes ($g(x) = 1/x$).
- Inverse-normal regression for positive, real-valued outcomes ($g(x) = 1/x^2$).

We consider the same data structure and assumptions as in Section 2.1 with the exception that the outcome Y is not restricted to be binary. We generalize the definitions of μ_0, μ_1 from Section 2.2 to be $\mu_0 = E(Y|A = 0), \mu_1 = E(Y|A = 1)$, respectively. We focus on testing the null hypothesis $H_0 : \mu_1 = \mu_0$.

The terms in the linear part of the GLM are assumed to be the same as in (1). Under such a generalized linear model with link function g, we have the following extension of (1):

$$g\{E(Y|A, B)\} = \gamma_0 + \gamma_1 A + \gamma_2' B. \tag{3}$$

If we further assume the link function is canonical, then it follows from Bickel and Doksum (2015, p. 413) that the corresponding maximum likelihood estimator $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2)$ for $(\gamma_0, \gamma_1, \gamma_2)$ is the solution to the following estimating equations:

$$\sum_{i=1}^{n} \left\{Y_i - g^{-1}(\gamma_0 + \gamma_1 A_i + \gamma_2' B_i)\right\} (1, A_i, B_i')' = \mathbf{0}. \tag{4}$$

In this section, we define $\hat{\gamma} = (\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2)$ to be the solution $(\gamma_0, \gamma_1, \gamma_2)$ to the above display. Define the GLM coefficient estimator to be $\hat{\gamma}_1$.

We consider the same setup and assumptions as in Section 3. Our results apply to a large class of smooth functions g, which include the five special cases given above. We make the following assumptions on g: g is continuously differentiable; g is invertible with strictly monotone inverse; for all γ we have $E\{\dot{h}(\gamma_0 + \gamma_1 A + \gamma_2' B)\}$ is finite, where $h = g^{-1}$ and $\dot{h}$ is the derivative of h. We also assume the regularity conditions of Theorems 5.9 and 5.21 of van der Vaart (1998, pp. 46, 52), which imply that $\hat{\gamma}$ converges in probability to the unique solution to $E\left\{Y - g^{-1}(\gamma_0 + \gamma_1 A + \gamma_2' B)\right\}(1, A, B')' = \mathbf{0}$, and is asymptotically normal. We do not assume that the model (3) is correctly specified.

For a given function g, the corresponding standardized estimator $\hat{\psi}_{std}$ is defined as (2) with expit replaced by g^{-1}, i.e.,

$$\hat{\psi}_{std} = \frac{1}{n}\sum_{i=1}^{n} g^{-1}(\hat{\gamma}_0 + \hat{\gamma}_1 + \hat{\gamma}_2' B_i) - \frac{1}{n}\sum_{i=1}^{n} g^{-1}(\hat{\gamma}_0 + \hat{\gamma}_2' B_i). \tag{5}$$

Rosenblum and van der Laan (2010) showed that the standardized estimator is a consistent estimator for the average treatment effect $E(Y|A=1) - E(Y|A=0)$ even if the model (3) is arbitrarily misspecified. They showed this for each of the five special cases of generalized linear models with canonical link functions g given above. We generalize their result by proving that it holds for the large class of functions g defined above. This result and a generalization of Theorem 2.1 are encapsulated in the following theorem:

Theorem 5.1. *Consider any function g satisfying the assumptions above. Then the standardized estimator $\hat{\psi}_{std}$ defined in (5) converges in probability to $E(Y|A=1) - E(Y|A=0)$. For each estimator $\hat{\psi}_{unadj}$, $\hat{\gamma}_1$, $\hat{\psi}_{std}$, it converges to 0 in probability if and only if H_0 is true. Therefore, the Wald test of H_0 based on any of these estimators has asymptotically correct Type I error. These results hold under arbitrary misspecification of the model (3).*

We next compare the power of $\hat{\gamma}_1$ and $\hat{\psi}_{std}$, generalizing Theorem 3.1.

Theorem 5.2. *Consider any function g satisfying the assumptions above. Let $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2)$ denote the solution to (4), and define the standardized estimator $\hat{\psi}_{std}$ as (5). Consider any set of alternatives $\{P^{(\nu)} : \nu > 0\}$ satisfying the assumptions in Section 3. The asymptotic relative efficiency for testing H_0, comparing Wald statistics based on $\hat{\psi}_{std}$ versus $\hat{\gamma}_1$, is 1.*

The theorem implies that the standardized estimator is asymptotically as efficient as $\hat{\gamma}_1$ for testing H_0. This holds under arbitrary misspecification of (3). Also, we prove a generalization of Corollary 1 to the setup of this section in the Appendix.

Consider the special case where $g(x) = x$. Then (3) is the linear regression model $E(Y|A, B) = \gamma_0 + \gamma_1 A + \gamma_2' B$. The GLM coefficient estimator $\hat{\gamma}_1$, defined as the solution to (4), is the ordinary least squares estimator of γ_1. Due to cancellation of terms in (5), the standardized estimator $\hat{\psi}_{std}$ is identical to $\hat{\gamma}_1$. This estimator is called the analysis of covariance (ANCOVA) estimator. Yang and Tsiatis (2001) proved that this estimator is consistent for the average treatment effect $E(Y|A=1) - E(Y|A=0)$ and has equal or greater precision compared to the unadjusted estimator, asymptotically, under arbitrary model misspecification. Theorem 5.2 holds trivially for the case of $g(x) = x$, since we have $\hat{\psi}_{std} = \hat{\gamma}_1$. This equality also holds for Poisson regression with canonical link function $g(x) = \log(x)$, if the standardized estimator uses the log rate-ratio contrast function $r(\mu_0, \mu_1) = \log(\mu_1/\mu_0)$. To the best of our knowledge, these are the only cases where $\hat{\gamma}_1$ and $\hat{\psi}_{std}$ are identical.

6 Discussion

Asymptotically valid confidence intervals for the average treatment effect can be constructed by using the nonparametric bootstrap applied to the standardized estimator. When baseline variables are moderately to strongly prognostic for the outcome, these confidence intervals can have shorter average widths than those constructed based on the unadjusted estimator, asymptotically.

The primary analysis in a confirmatory randomized trial needs to be prespecified in the study protocol. When using an adjusted estimator, this requires specifying the list of baseline variables to be used and the precise method (e.g., the standardized estimator using a logistic regression model with main terms for treatment and baseline variables). A challenging practical problem is how to select baseline variables. The number of variables should not be too large compared to the sample size of the trial, though it is an open problem to determine what "too large" means. A conservative approach would be to pick a few key baseline variables that are expected to be prognostic for the outcome based on clinical knowledge and prior data.

We focused on trials with simple randomization, but randomization stratified on key covariates can also be used. In the latter case, the standardized estimator can adjust for the stratification variables and additional baseline variables. Stratified randomization can typically only be applied to at most a few binary-valued variables (since otherwise some stratum becomes too small to balance by design). This may leave other prognostic variables that can be adjusted for to improve precision by using an adjusted estimator.

The standardized estimator can be modified to handle missing outcome data, as described, e.g., by Moore and van der Laan (2009); Colantuoni and Rosenblum (2015). The estimator is consistent for the average treatment effect under the following assumptions: the missing at random assumption, a correctly modeled probability of censoring given baseline variables and study arm, and the assumption that this probability is bounded away from 1.

The standardized estimator for binary outcomes has been applied in simulation studies by, e.g., Moore and van der Laan (2009); Colantuoni and Rosenblum (2015); Steingrimsson et al. (2016), where substantial efficiency gains were observed compared to the unadjusted estimator. Steingrimsson et al. (2016) provide

R and Stata code that implement the standardized estimator and compute confidence intervals based on the nonparametric bootstrap.

We focused on the standardized estimator (2) for its simplicity and ease of computation. Our results have implications for the enhanced efficiency, standardized estimators of, e.g., Robins et al. (2007); Tan (2010); Rotnitzky et al. (2012); Gruber and van der Laan (2012); Colantuoni and Rosenblum (2015), all of which have equal or better efficiency compared to the standardized estimator (2) in our context of a randomized trial (under suitable regularity conditions). Compared to (2), these estimators are more complex (and some are more computationally challenging), but they have potential for greater efficiency gains when the logistic regression model is misspecified. Colantuoni and Rosenblum (2015) describe these tradeoffs.

7 Acknowledgments

We sincerely thank Daniel Hanley and the MISTIE team.

Appendix

We start by proving Theorem 5.1, which is used to prove several of the other results.

Proof of Theorem 5.1. Assume the conditions in Section 5. Let $\hat{\gamma}$ denote the solution to the estimating equations (4). For each $a \in \{0, 1\}$, define

$$\hat{\psi}^{(a)}_{std} = \frac{1}{n} \sum_{i=1}^{n} g^{-1}(\hat{\gamma}_0 + \hat{\gamma}_1 a + \hat{\gamma}_2' B_i). \tag{6}$$

We will prove that $\hat{\psi}^{(a)}_{std}$ converges in probability to $E(Y|A = a)$, for each $a \in \{0, 1\}$. (Throughout this proof, expectation E is with respect to P_0.)

Since $\hat{\gamma}$ is the solution to the estimating equations (4), it follows from the regularity conditions (which we assumed in Section 5) of Theorems 5.9 and 5.21 of van der Vaart (1998, pp. 46, 52) that $\hat{\gamma}$ converges in probability to the solution $\bar{\gamma}$ to $E\left\{Y - g^{-1}(\gamma_0 + \gamma_1 A + \gamma_2' B)\right\}(1, A, B')' = \mathbf{0}$. This implies

$$\begin{aligned} EY &= Eg^{-1}(\bar{\gamma}_0 + \bar{\gamma}_1 A + \bar{\gamma}_2' B) = (1/2)Eg^{-1}(\bar{\gamma}_0 + \bar{\gamma}_2' B) + (1/2)Eg^{-1}(\bar{\gamma}_0 + \bar{\gamma}_1 + \bar{\gamma}_2' B); \\ EAY &= EAg^{-1}(\bar{\gamma}_0 + \bar{\gamma}_1 A + \bar{\gamma}_2' B) = (1/2)Eg^{-1}(\bar{\gamma}_0 + \bar{\gamma}_1 + \bar{\gamma}_2' B), \end{aligned}$$

which follow from A and B being independent. Since $EY = EYA + EY(1 - A)$, it follows that for each $a \in \{0, 1\}$, we have $E(Y|A = a) = Eg^{-1}(\bar{\gamma}_0 + \bar{\gamma}_1 a + \bar{\gamma}_2' B)$. Since for each $a \in \{0, 1\}$ the estimator $\hat{\psi}^{(a)}_{std}$ converges in probability to $Eg^{-1}(\bar{\gamma}_0 + \bar{\gamma}_1 a + \bar{\gamma}_2' B)$, we have shown $\hat{\psi}^{(a)}_{std}$ converges in probability to $E(Y|A = a)$. This shows the standardized estimator $\hat{\psi}_{std} = \hat{\psi}^{(1)}_{std} - \hat{\psi}^{(0)}_{std}$ converges in probability to $E(Y|A = a) - E(Y|A = 0)$. This also holds for $\hat{\psi}_{unadj}$. Therefore, each of the estimators converges to 0 in probability if and only if H_0 is true. It remains to show this for $\hat{\gamma}_1$.

It follows from the above arguments that

$$E(Y|A=1)-E(Y|A=0)=E\{g^{-1}(\bar\gamma_0+\bar\gamma_1+\bar\gamma_2'B)\}-E\{g^{-1}(\bar\gamma_0+\bar\gamma_2'B)\}. \tag{7}$$

Since we assumed $g^{-1}(x)$ is strictly monotone, the right side of (7) equals 0 if and only if $\bar\gamma_1=0$. Therefore, $\hat\gamma_1$ converges to 0 in probability if and only if H_0 is true. □

Proof of Theorem 5.2. The estimators below are as defined in Section 5. By Theorem 5.1, both the GLM coefficient estimator and the standardized estimator converge to 0 in probability under any $P^{(0)}\in H_0$. We use the following change of variables: $A^*=2A-1$. Then (3) is equivalent to the following:

$$E(Y|A^*,B)=g^{-1}(\gamma_0^*+\gamma_1^*A^*+\gamma_2^{*\prime}B), \tag{8}$$

where $\gamma_0^*=\gamma_0+\gamma_1/2$, $\gamma_1^*=\gamma_1/2$, $\gamma_2^*=\gamma_2$. Denote the corresponding estimated coefficients by adding the hat symbol. The asymptotic relative efficiency is unchanged if we substitute A^* for A and $\hat\gamma_1^*$ for $\hat\gamma_1$, respectively, which we do below.

We assumed the regularity conditions in Lemma 7.6 of van der Vaart (1998, pp. 95), which include the following: the set of distributions $\{P^{(\nu)}:\nu\geq 0\}$ is dominated by a common measure λ, with corresponding probability densities $p^{(\nu)}$; the map $\nu\mapsto\{p^{(\nu)}\}^{1/2}$ is continuously differentiable for every point in the sample space, and all components of $\int\{(dp^{(\nu)}/d\nu)/p^{(\nu)}\}^2p^{(\nu)}d\lambda$ are well defined and the integral is continuous in ν. Lemma 7.6 implies that $\{P^{(\nu)}:\nu\geq 0\}$ is differentiable in quadratic mean with score function $\dot l^{(\nu)}(B,A^*,Y)=1_{p^{(\nu)}>0}(dp^{(\nu)}/d\nu)/p^{(\nu)}$, where $1_{p^{(\nu)}>0}$ is the indicator variable taking value 1 if $p^{(\nu)}>0$ and 0 otherwise. Differentiability in quadratic mean at $\nu=0$ is assumed to be from the right, i.e., $\int\left[\{p^{(\nu)}\}^{1/2}-\{p^{(0)}\}^{1/2}-(1/2)\nu\dot l^{(0)}\{p^{(0)}\}^{1/2}\right]^2d\lambda=o(\nu^2)$ as $\nu\downarrow 0$. We let $\dot l(B,A^*,Y)$ with no superscript denote $\dot l^{(0)}(B,A^*,Y)$.

Define $\bar\gamma^*(\nu)$ to be the probability limit (as $n\to\infty$) under $P^{(\nu)}$ of the estimator $\hat\gamma^*$. Let $IF_{GLM}(B,A^*,Y)$ and $IF_{std}(B,A^*,Y)$ denote the influence functions for the estimators $\hat\gamma_1^*$ and $\hat\psi_{std}$, respectively, under $P^{(0)}$. It follows from Theorem 5.21 of van der Vaart (1998, p. 52) that

$$IF_{GLM}(B,A^*,Y) = \frac{1}{E^{(0)}\dot h\left\{\bar\gamma_0^*(0)+\bar\gamma_2^{*\prime}(0)B\right\}}\left[Y-g^{-1}\left\{\bar\gamma_0^*(0)+\bar\gamma_2^{*\prime}(0)B\right\}\right]A^*, \tag{9}$$

$$IF_{std}(B,A^*,Y) = 2\left[Y-g^{-1}\left\{\bar\gamma_0^*(0)+\bar\gamma_2^{*\prime}(0)B\right\}\right]A^*, \tag{10}$$

where we used that $\bar\gamma_1^*(0)=0$ under $P^{(0)}$, which follows from Theorem 5.1. Define $\mathrm{avar}_{GLM}(\nu)$ and $\mathrm{avar}_{std}(\nu)$ to be the variance of $IF_{GLM}(B,A^*,Y)$ and $IF_{std}(B,A^*,Y)$, respectively, under $P^{(\nu)}$. Direct calculation gives that at $\nu=0$,

$$\mathrm{avar}_{GLM}(0) = \left\{E^{(0)}\dot h\left(\bar\gamma_0^*+\bar\gamma_2^{*\prime}B\right)\right\}^{-2}E^{(0)}\left[\left\{Y-h(\bar\gamma_0^*+\bar\gamma_2^{*\prime}B)\right\}^2\right], \tag{11}$$

$$\mathrm{avar}_{std}(0) = 4E^{(0)}\left[\left\{Y-h(\bar\gamma_0^*+\bar\gamma_2^{*\prime}B)\right\}^2\right]. \tag{12}$$

By the above conditions and the assumption that $P^{(\nu)}$ converges in total variation distance to $P^{(0)}$ as $\nu\downarrow 0$, it follows that for any bounded, measurable function $f(B,A^*,Y)$, we have

$$E^{(0)}\dot l(B,A^*,Y)f(B,A^*,Y)=-\frac{d}{d\nu}E^{(\nu)}f(B,A^*,Y)\bigg|_{\nu=0^+}, \tag{13}$$

where the $+$ indicates the derivative is taken from the right. Also, since A^* is independent of B for each $P^{(\nu)}$, the score $\dot{l}(B, A^*, Y)$ is orthogonal (under $E^{(0)}$) to any square-integrable function of A^*, B that has mean zero given B. We then have

$$\begin{aligned} E^{(0)}\dot{l}(B, A^*, Y)IF_{std}(B, A^*, Y) &= 2E^{(0)}\dot{l}(B, A^*, Y)\left[Y - g^{-1}\left\{\bar{\gamma}_0^*(0) + \bar{\gamma}_2^{*\prime}(0)B\right\}\right]A^* \\ &= 2E^{(0)}\dot{l}(B, A^*, Y)Y \quad (14) \\ &= -\frac{d}{d\nu}\left(2E^{(\nu)}YA^*\right)\Big|_{\nu=0^+} \quad (15) \\ &= -\frac{d}{d\nu}\left\{\mu_1(\nu) - \mu_0(\nu)\right\}\Big|_{\nu=0^+} < 0, \quad (16) \end{aligned}$$

where (14) follows from $E^{(0)}(A^*|B) = E^{(0)}A^* = 0$ and that $\dot{l}(B, A^*, Y)$ is orthogonal (under $E^{(0)}$) to any square-integrable function of A^*, B that has mean zero given B; (15) follows from (13); (16) follows from the assumption that $\mu_1(\nu) - \mu_0(\nu)$ is right differentiable with positive right-derivative. It follows from analogous arguments as above, but applied to IF_{GLM} instead of IF_{std} and using (9), that

$$E^{(0)}\dot{l}(B, A^*, Y)IF_{GLM}(B, A^*, Y) = -\frac{1}{2E^{(0)}\dot{h}\left\{\bar{\gamma}_0^*(0) + \bar{\gamma}_2^{*\prime}(0)B\right\}}\frac{d}{d\nu}\left\{\mu_1(\nu) - \mu_0(\nu)\right\}\Big|_{\nu=0^+} \quad (17)$$

We next apply Le Cam's Third Lemma (van der Vaart, 1998, p. 90) to derive the asymptotic distributions of the two estimators under local alternatives $P^{(\nu_n)}$ for ν_n proportional to $n^{-1/2}$, as $n \to \infty$. The conditions of the lemma are met since our assumed regularity conditions imply local asymptotic normality (van der Vaart, 1998, p. 94), and mutual contiguity of $P^{(\nu)}$ and $P^{(0)}$ (as $\nu \downarrow 0$) holds by our assumption that $P^{(\nu)}$ converges to $P^{(0)}$ in total variation distance. Define

$$\underline{\gamma}_1^*(\nu) = \nu E^{(0)}\dot{l}(B, A^*, Y)IF_{GLM}(B, A^*, Y), \qquad \underline{\psi}(\nu) = \nu E^{(0)}\dot{l}(B, A^*, Y)IF_{std}(B, A^*, Y). \quad (18)$$

Le Cam's Third Lemma implies for any $h \geq 0$, for $\nu_n = hn^{-1/2}$,

$$n^{1/2}\left\{\frac{\hat{\gamma}_1^* - \underline{\gamma}_1^*(\nu_n)}{\text{avar}_{GLM}(\nu_n)^{1/2}}\right\} \xrightarrow{\nu_n} N(0,1), \qquad n^{1/2}\left\{\frac{\hat{\psi}_{std} - \underline{\psi}(\nu_n)}{\text{avar}_{std}(\nu_n)^{1/2}}\right\} \xrightarrow{\nu_n} N(0,1), \quad (19)$$

where for each $n > 0$, the estimators $\hat{\gamma}_1$ and $\hat{\psi}_{std}$ in the expressions above are based on n independent, identically distributed draws from $P^{(\nu_n)}$.

Define $\dot{\gamma}_1^* = d\underline{\gamma}_1^*/d\nu\Big|_{\nu=0^+}$ and $\dot{\psi} = d\underline{\psi}/d\nu\big|_{\nu=0^+}$. It follows from Theorem 14.19 of van der Vaart (1998, p. 201) that the asymptotic relative efficiency comparing the standardized and GLM coefficient estimators equals the following square of the ratio of slopes of the corresponding statistics:

$$\left\{\frac{\dot{\gamma}_1^*}{\dot{\psi}}\right\}^2\left\{\frac{\text{avar}_{GLM}(0)}{\text{avar}_{std}(0)}\right\}^{-1}. \quad (20)$$

It follows from (16), (17), and (18) that $\dot{\gamma}_1^*/\dot{\psi} = \left[2E^{(0)}\dot{h}\left\{\bar{\gamma}_0^*(0) + \bar{\gamma}_2^{*\prime}(0)B\right\}\right]^{-1}$. It then follows from (11), (12), and (20) that the asymptotic relative efficiency of the GLM estimator $\hat{\gamma}_1^*$ versus the standardized estimator is

$$\left[\frac{1}{2E\dot{h}\left\{\bar{\gamma}_0^*(0)+\bar{\gamma}_2^{*\prime}(0)B\right\}}\right]^2\left(\frac{4E\left[\left\{Y-h(\bar{\gamma}_0^*(0)+\bar{\gamma}_2^{*\prime}(0)B)\right\}^2\right]}{\left\{E\dot{h}\left(\bar{\gamma}_0^*(0)+\bar{\gamma}_2^{*\prime}(0)B\right)\right\}^{-2}E\left[\left\{Y-h(\bar{\gamma}_0^*(0)+\bar{\gamma}_2^{*\prime}(0)B)\right\}^2\right]}\right)=1,$$

where all expectations are with respect to $P^{(0)}$. This completes the proof of Theorem 5.2. $\square$

Theorems 2.1 and 3.1 follow from analogous arguments as Theorems 5.1 and 5.2, respectively, using $g(x) = \text{logit}(x)$. In this special case, $\hat{\gamma}$ converges to the maximizer of the expected log-likelihood given in Section 2.3, and asymptotic normality of $\hat{\gamma}$ follows from strict concavity of the expected log-likelihood and Theorem 5.23 of van der Vaart (1998, p. 53).

Proof of Corollary 3.2. We prove the generalization of Corollary 1 to the setup in Section 5. Consider any function g satisfying the assumptions in that section. The asymptotic relative efficiency of the GLM coefficient estimator compared to the unadjusted estimator equals the analog of (20) with $\hat{\psi}_{std}$ replaced throughout by $\hat{\psi}_{unadj}$. The first term in curly braces is unchanged, since $\hat{\psi}_{unadj}$ and $\hat{\psi}_{std}$ are consistent estimators of the same quantity. The second term in curly braces involves replacing avar$_{std}$ by avar$_{unadj}$. If the model (3) is correctly specified, we have avar$_{std}(0) \leq$ avar$_{unadj}(0)$. Therefore, the asymptotic relative efficiency of the GLM coefficient estimator $\hat{\bar{\gamma}}_1^*$ compared to the unadjusted estimator $\hat{\psi}_{unadj}$ is at least 1. Combining this result with Theorem 5.2 implies the generalization of Corollary 1 to the setup in Section 5. $\square$

Proof of Theorem 3.3. Let $\nabla r(x,y) = [\dot{r}_1(x,y), \dot{r}_2(x,y)]'$, where the first and second component denote the partial derivative w.r.t. the first and second component of $r(x,y)$, respectively. The condition that $r = 0$ under H_0 implies that $r(x,x) = 0$ for all x. Differentiating both sides with respect to x implies $\dot{r}_1(x,x) + \dot{r}_2(x,x) = 0$, from which it follows that

$$\nabla r(x,x) = [\dot{r}_1(x,x), -\dot{r}_1(x,x)]' = \dot{r}_1(x,x)[1,-1]'. \tag{21}$$

When using the contrast r, the parameter estimated by the standardized estimator under $P^{(\nu)}$ is denoted $\psi_{std,r}(\nu) = r(\mu_0(\nu), \mu_1(\nu))$. By our assumption that $P^{(0)} \in H_0$, we have $\mu_0(0) = \mu_1(0)$. It follows from equation (21) and the chain rule that

$$\begin{aligned}\left.\frac{d\psi_{std,r}}{d\nu}\right|_{\nu=0^+} &= \nabla r(\mu_0(0),\mu_1(0))' * [\partial\mu_0/\partial\nu, \partial\mu_1/\partial\nu]'|_{\nu=0^+} \\ &= \dot{r}_1(\mu_0(0),\mu_0(0))[\partial\mu_0/\partial\nu - \partial\mu_1/\partial\nu]|_{\nu=0^+}. \end{aligned} \tag{22}$$

The delta method and (21) imply that the asymptotic variance of the standardized estimator using the contrast function r at $\nu = 0$ is avar$_{std,r}(0) =$ avar$_{std}(0)[\dot{r}_1(\mu_0(0),\mu_0(0))]^2$, where avar$_{std}(0)$ is defined in equation (12). Combining this with (22) shows that

$$\left(\left.\frac{d\psi_{std,r}}{d\nu}\right|_{\nu=0^+}\right)^2 / \text{avar}_{std,r}(0) = \frac{[(\partial\mu_0/\partial\nu - \partial\mu_1/\partial\nu)|_{\nu=0^+}]^2}{\text{avar}_{std}(0)},$$

which is independent of the choice of r. This combined with equation (20) proves Theorem 3.3. $\square$

References

Peter C Austin, Andrea Manca, Merrick Zwarenstein, David N Juurlink, and Matthew B Stanbrook. A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals. *Journal of Clinical Epidemiology*, 63(2): 142–153, 2010.

Peter J Bickel and Kjell A Doksum. *Mathematical statistics: basic ideas and selected topics*, volume 1. Chapman and Hall, CRC Press, Boca Raton, 2015.

Elizabeth Colantuoni and Michael Rosenblum. Leveraging prognostic baseline variables to gain precision in randomized trials. *Statistics in Medicine*, 34(18):2602–2617, 2015.

David A Freedman. Randomization does not justify logistic regression. *Statistical Science*, 23(2):237–249, 2008.

Mitchell H Gail, S Wieand, and Steven Piantadosi. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*, 71(3):431–444, 1984.

S. Gruber and MJ. van der Laan. Targeted minimum loss based estimator that outperforms a given estimator. *The International Journal of Biostatistics*, 8(1), 2012.

D. F. Hanley, R. E. Thompson, J. Muschelli, M. Rosenblum, N. McBee, K. Lane, A. J. Bistran-Hall, S. W. Mayo, P. Keyl, T. C. Morgan, N. Ullman, W. A. Mould, J. R. Carhuapoma, C. Kase, W. Ziai, C. B. Thompson, G. Yenokyan, E. Huang, W. C. Broaddus, R. S. Graham, E. F. Aldrich, R. Dodd, C. Wijman, J-L. Caron, J. Huang, P. Camarata, S. Janis, P. Vespa, N. Martin, I. Awad, and Zuccarello M. for the MISTIE II Investigators. Safety and efficacy of minimally invasive surgery plus recombinant tissue plasminogen activator in intracerebral haemorrhage evacuation (mistie): a randomised phase 2 trial. *Lancet Neurology*, In Press.

Adrián V Hernández, Ewout W Steyerberg, Isabella Butcher, Nino Mushkudiani, Gillian S Taylor, Gordon D Murray, Anthony Marmarou, Sung C Choi, Juan Lu, J Dik F Habbema, et al. Adjustment for strong predictors of outcome in traumatic brain injury trials: 25% reduction in sample size requirements in the IMPACT study. *Journal of Neurotrauma*, 23(9):1295–1303, 2006.

Peter McCullagh and John A Nelder. *Generalized linear models*, volume 37. Chapman and Hall, CRC press, Boca Raton, 1989.

Kelly L Moore and Mark J van der Laan. Covariate adjustment in randomized trials with binary outcomes: targeted maximum likelihood estimation. *Statistics in Medicine*, 28(1):39, 2009.

JM. Robins, M. Sued, Q. Lei-Gomez, and A. Rotnitzky. Double-robustness with improved efficiency in missing and causal inference models. *Technical report, Harvard School of Public Health*, 2007.

Laurence D Robinson and Nicholas P Jewell. Some surprising results about covariate adjustment in logistic regression models. *International Statistical Review/Revue Internationale de Statistique*, pages 227–240, 1991.

M. Rosenblum and MJ. van der Laan. Simple, efficient estimators of treatment effects in randomized trials using generalized linear models to leverage baseline variables. *International Journal of Biostatistics*, 6 (1):Article 1, 2010.

Michael Rosenblum and Mark J. van der Laan. Using regression models to analyze randomized trials: Asymptotically valid hypothesis tests despite incorrectly specified models. *Biometrics*, 65(3):937–945, 2009. ISSN 1541-0420. doi: 10.1111/j.1541-0420.2008.01177.x. URL `http://dx.doi.org/10.1111/j.1541-0420.2008.01177.x`.

A. Rotnitzky, Q. Lei, M. Sued, and JM. Robins. Improved double-robust estimation in missing data and causal inference models. *Biometrika*, 99(2):439–456, 2012.

Daniel O Scharfstein, Andrea Rotnitzky, and James M Robins. Rejoinder to "adjusting for nonignorable drop-out using semiparametric nonresponse models". *Journal of the American Statistical Association*, 94 (448):1096–1120, 1999.

Jon Arni Steingrimsson, Daniel F. Hanley, and Michael Rosenblum. Improving precision by adjusting for baseline variables in randomized trials with binary outcomes, without regression model assumptions. *Johns Hopkins University, Dept. of Biostatistics Working Papers. http://biostats.bepress.com/jhubiostat/paper280*, 2016.

Z. Tan. Bounded, efficient and doubly robust estimating equations for marginal and nested structural models. *Biometrika*, 97:661–82, 2010.

AW. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998.

Li Yang and Anastasios A Tsiatis. Efficiency study of estimators for a treatment effect in a pretest–posttest trial. *The American Statistician*, 55(4):314–321, 2001.