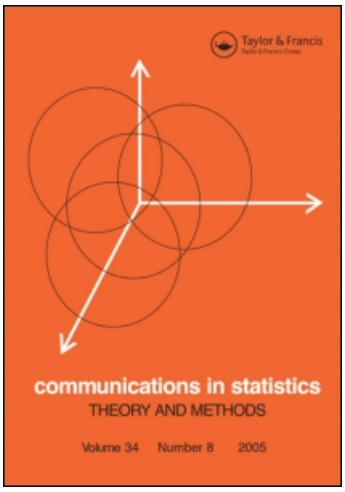

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597238

Estimating a Marginal Causal Odds Ratio Subject to Confounding

Zhiwei Zhang [a]
[a] Center for Devices and Radiological Health, U.S. Food and Drug Administration, Rockville, Maryland, USA

Online Publication Date: 01 January 2009

To cite this Article Zhang, Zhiwei(2009)'Estimating a Marginal Causal Odds Ratio Subject to Confounding',Communications in Statistics - Theory and Methods,38:3,309 — 321
To link to this Article: DOI: 10.1080/03610920802200076
URL: http://dx.doi.org/10.1080/03610920802200076

Communications in Statistics—Theory and Methods, 38: 309–321, 2009

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920802200076

Taylor & Francis
Taylor & Francis Group

Estimating a Marginal Causal Odds Ratio Subject to Confounding

ZHIWEI ZHANG

Center for Devices and Radiological Health, U.S. Food and Drug Administration, Rockville, Maryland, USA

Odds ratios are frequently used to describe the relationship between a binary treatment or exposure and a binary outcome. An odds ratio can be interpreted as a causal effect or a measure of association, depending on whether it involves potential outcomes or the actual outcome. An odds ratio can also be characterized as marginal versus conditional, depending on whether it involves conditioning on covariates. This article proposes a method for estimating a marginal causal odds ratio subject to confounding. The proposed method is based on a logistic regression model relating the outcome to the treatment indicator and potential confounders. Simulation results show that the proposed method performs reasonably well in moderate-sized samples and may even offer an efficiency gain over the direct method based on the sample odds ratio in the absence of confounding. The method is illustrated with a real example concerning coronary heart disease.

Keywords Causal inference; Collapsibility; Confounding; Logistic regression; Odds ratio.

Mathematics Subject Classification Primary 62F10, 62F12; Secondary 62P10, 62H12.

1. Introduction

Odds ratios (ORs) are frequently used to describe the relationship between a binary treatment or exposure and a binary outcome. The interpretation of an OR depends on its precise definition. An OR defined in terms of the conditional distribution of the actual outcome given treatment assignment (and possibly other variables) is a measure of association. An OR defined in terms of the (conditional) distributions of potential outcomes under different treatments is a causal effect. For clarity, the former type of ORs will be said to be associational and the latter causal. When an associational OR differs from its causal counterpart, which occurs quite often in observational studies, the former quantity is said to be confounded for the latter (Greenland et al., 1999).

Received September 11, 2007; Accepted May 13, 2008
Address correspondence to Zhiwei Zhang, Division of Biostatistics, OSB/CDRH/FDA, HFZ-550, Rockville, MD 20850, USA; E-mail: zhiwei.zhang@fda.hhs.gov

An OR may also be described as marginal versus conditional, depending on whether it involves conditioning on variables other than treatment assignment. In modern clinical and epidemiological studies, information is often collected on a number of covariates that may be related to treatment assignment and/or potential outcomes. The availability of covariate information makes it possible to estimate conditional ORs, often through a logistic regression analysis. Conditional ORs are generally different from their marginal counterparts. This phenomenon has been noted by Gail et al. (1984) in the context of generalized linear models and is also known as non-collapsibility in causal inference (Greenland et al., 1999).

There is a lot of literature on causal inference in observational studies (e.g., Bang and Robins, 2005; D'Agostino, 1998; Lunceford and Davidian, 2004). Yet, to my knowledge, little discussion has been devoted to the problem of estimating a marginal causal odds ratio (MCOR) in the presence of confounding. It is well known that the sample OR, which ignores the covariate information, is generally biased for the MCOR when treatment assignment depends on potential outcomes. A logistic regression model is often used to adjust for potential confounders. The regression coefficients in a logistic regression model have a causal interpretation if all important confounders are included and the model specification is approximately correct. However, the regression coefficients are more closely related to conditional ORs than to the MCOR. Recently, Austin (2007) compared several estimators based on propensity scores in simulation experiments and found all of them to be more or less biased for the MCOR. I am not aware of an existing estimator that has been demonstrated to be consistent for the MCOR in the presence of confounding. This article proposes a consistent estimator of the MCOR based on a standard logistic regression model for the outcome given treatment assignment and covariates. While the proposed estimator is designed primarily for observational studies, it can certainly be used in randomized studies and may offer an efficiency gain over the sample OR, as will be shown in a simulation study.

The rest of the article is organized as follows. Section 2 sets up the notation and terminology. Section 3 describes the methodology. Section 4 presents numerical results, and Sec. 5 concludes the article with a discussion.

2. Notation and Terminology

Let us begin by formulating the statistical problem in the potential outcome notation. For a subject chosen randomly from a given population, let $Y(t)$ denote the (potential) outcome that will realize if the subject receives the standard treatment ($t = 0$) or a new treatment ($t = 1$). The outcome is assumed binary, with 0 corresponding to failure and 1 to success. The actual treatment assignment will be denoted by $T = 0, 1$, and the actual outcome by $Y = Y(T)$. Let X represent a collection of covariates.

There are different ways to define an OR in the present setting, four of which are shown in Table 1. For ease of display, an OR in Table 1 is denoted by a pair of probabilities to be compared and not by the entire formula that actually yields the OR; thus an entry of the form $p_0 : p_1$ represents an OR given by $(1 - p_0)p_1/\{p_0(1 - p_1)\}$. The entries of Table 1 can be categorized by interpretation as associational versus causal. Generally, ORs defined in terms of the conditional distribution of Y given T and possibly X measure the (conditional) association between T and Y, while those defined in terms of the (conditional) distributions of $Y(0)$ and $Y(1)$

Table 1
Four different ways to define an odds ratio

	Associational	Causal
Marginal	$P[Y = 1 \mid T = 0]$: $P[Y = 1 \mid T = 1]$	$P[Y(0) = 1]$: $P[Y(1) = 1]$
Conditional	$P[Y = 1 \mid T = 0, X]$: $P[Y = 1 \mid T = 1, X]$	$P[Y(0) = 1 \mid X]$: $P[Y(1) = 1 \mid X]$

represent causal effects of the new treatment relative to the standard treatment. The entries of Table 1 can also be classified as marginal or conditional, depending on whether the definition involves conditioning on X. The MCOR can be used to assess the overall effect of the new treatment on the entire population, while the conditional causal OR may help a clinician decide which medical treatment to recommend, given relevant information about a particular patient. Marginal ORs are unique in a given population, while conditional ORs are defined with respect to a set of covariates. In practice, characterization of a conditional OR often relies on correct specification of a logistic regression model and may be complicated by the presence of interactions. The focus of this article is on estimation of the MCOR.

Table 1 illustrates simultaneously the distinct issues of confounding and collapsibility, both discussed in depth by Greenland et al. (1999). In the present context, confounding refers to any difference between a causal OR and its associational counterpart. For example, the marginal associational OR may be confounded for the MCOR if treatment assignment depends on potential outcomes. Confounding is parameter-specific, so it may be that the marginal OR is confounded but the conditional is not. Collapsibility means here that a conditional OR, causal or associational, is independent of the conditioning variable(s) and identical to its marginal counterpart. ORs are not in general collapsible, a phenomenon known as Simpson's paradox. In fact, the issue of non collapsibility arises in any generalized linear model with a link function that is not identity or a simple log (Gail et al., 1984). In a very special case, the causal ORs are collapsible if potential outcomes are independent of covariates so that conditioning on covariates does not alter the success probabilities of potential outcomes. This article deals with the more general setting with confounding and non collapsibility both present for the MCOR.

3. Methodology

We assume a prospective design and suppose the observed data can be represented as (X_i, Y_i, T_i), $i = 1, \ldots, n$, which are independent and identically distributed as (X, Y, T). Identification of the MCOR, or any causal parameter, generally requires assumptions. It is assumed here that given covariates, treatment assignment is conditionally independent of the potential outcomes, written

$$T \perp \{Y(0), Y(1)\} \mid X. \tag{1}$$

Heuristically, this assumption requires that X fully explains the association between T and the potential outcomes so there are no unmeasured confounders. Assumption (1) implies that the conditional ORs are unconfounded, because

$P[Y = 1 \mid T = t, X] = P[Y(t) = 1 \mid T = t, X] = P[Y(t) = 1 \mid X]$, $t = 0, 1$. The MCOR, on the other hand, is still subject to confounding.

The main idea of this article is best explained by first showing how biased estimates may result from standard methods. An obvious candidate is the sample OR based on the (Y_i, T_i) only, which estimates the marginal associational OR and may be biased for the MCOR due to confounding. This bias can in principle be corrected by utilizing information in the measured confounders and by invoking the assumption of no unmeasured confounders. A simple way to adjust for confounders would be a logistic regression analysis where Y is the response variable and X and T are predictors. The regression coefficients in such a model relate directly to the (unconfounded) conditional OR. For example, if there is no interaction between T and X, then the conditional OR is constant and its natural log is given by the regression coefficient for T. While the conditional OR is unconfounded, there is now the issue of non collapsibility: The MCOR does not derive from its conditional counterpart in a simple manner.

There is also a class of methods based on the notion of propensity score. The propensity score is the conditional probability, given covariates, of receiving the new treatment. Given the propensity score, treatment assignment is conditionally independent of the potential outcomes, so that a fair comparison can be made at each level of the propensity score (Rosenbaum and Rubin, 1983). Austin (2007) compared several propensity score methods through simulations and found all of them to be more or less biased, at least in some settings, for the MCOR. It is impractical to analyze all existing propensity score methods (e.g., matching, stratification, regression adjustment, inverse probability weighting) under each implementation in the present article. Nonetheless it seems instructive to illustrate a common pitfall using as an example a simple stratification procedure examined by Austin et al. (2007) and Austin (2007). This procedure basically stratifies the sample according to estimated propensity scores, obtains an OR estimate from each stratum, and takes the average of the stratum-specific OR estimates. Stratification by the estimated propensity score can be quite effective for bias reduction in the presence of confounding (Rosenbaum and Rubin, 1984), so a stratum-specific OR may have a causal interpretation, at least approximately. On the other hand, stratification also raises the issue of non collapsibility. A stratum-specific OR is really a conditional OR, conditioning on the event that the propensity score falls into some range. Averaging over the conditional ORs does not necessarily yield the MCOR.

The foregoing discussion suggests that one should be mindful of non collapsibility while dealing with confounding. Although adjustments for confounders often make it convenient to work with conditional ORs, it appears difficult to derive an estimate of the MCOR from estimates of conditional ORs. Indeed, knowledge of conditional ORs is neither sufficient nor necessary for determining the MCOR. The approach of this article is to bypass the conditional ORs and work instead with the marginal success probabilities

$$\pi_t = P[Y(t) = 1], \quad t = 0, 1.$$

The log-MCOR is then given by

$$\phi = \text{logit}(\pi_1) - \text{logit}(\pi_0), \tag{2}$$

where $\text{logit}(p) = \log\{p/(1-p)\}$. No treatment comparisons will be made until estimates of the π_t are plugged in to yield an estimate of ϕ. Estimation of each π_t can be regarded as a missing data problem, since the potential outcome $Y(t)$ is observed only for subjects who actually receive treatment t. Assumption (1) implies that the missing outcomes are missing at random in the sense of Rubin (1976). A systematic discussion of missing data can be found in Little and Rubin (2002). All available methods for estimating π_t will not be discussed here. In what follows, I describe a simple imputation-type method to illustrate how the general idea might be implemented.

The method is motivated by the observation that

$$\begin{aligned}\pi_t &= E[P[Y(t) = 1 \mid X]] = E[P[Y(t) = 1 \mid T = t, X]] \\ &= E[P[Y = 1 \mid T = t, X]], \quad t = 0, 1,\end{aligned} \tag{3}$$

which follows from assumption (1). The above representation shows that the marginal distributions of the potential outcomes, as characterized by the π_t, can be "synthesized" from the regression of Y on (T, X) and the covariate distribution. Suppose that given (T, X), Y follows a logistic regression model:

$$\text{logit}(P[Y = 1 \mid T, X]) = h(T, X; \beta), \tag{4}$$

where h is a known function and β an unknown, finite-dimensional parameter. This is known as the outcome regression model (Bang and Robins, 2005) because the response variable in the model is the observed outcome as opposed to treatment assignment. It would be computationally convenient to have $h(T, X; \beta)$ linear in β; this is not really restrictive because $h(T, X; \beta)$ can involve arbitrary transformations of X and/or interactions with T. The exact form of h may be difficult to specify. However, unless X has only one or two continuous components, it may be unrealistic to rely on nonparametric regression techniques due to the "curse of dimensionality" (Robins and Ritov, 1997). It is clear from (3) that model (4) is equivalent to a pair of models for the two potential outcomes:

$$\text{logit}(P[Y(t) = 1 \mid X]) = h(t, X; \beta), \quad t = 0, 1,$$

which may or may not take the same form or share any parameters.

Let $\hat{\beta}$ denote the maximum likelihood estimate of β, which can be found by solving the likelihood equations:

$$\sum_{i=1}^{n} \dot{h}(T_i, X_i; \beta)[Y_i - \text{logit}^{-1}\{h(T_i, X_i; \beta)\}] = 0,$$

where $\dot{h}(t, x; \beta) = \partial h(t, x; \beta)/\partial\beta$. Now expression (3) suggests that the π_t be estimated by

$$\hat{\pi}_t = \frac{1}{n}\sum_{i=1}^{n} \text{logit}^{-1}\{h(t, X_i; \hat{\beta})\}, \quad t = 0, 1.$$

This can be regarded as an imputation method that replaces each potential outcome, observed or not, by its predicted value based on X_i. Substituting the above estimates into (2) leads to

$$\hat{\phi} = \text{logit}(\hat{\pi}_1) - \text{logit}(\hat{\pi}_0).$$

As will be shown in Appendix A, under model (4) and regularity conditions, $\sqrt{n}(\hat{\phi} - \phi) \to N(0, \sigma_1^2 + \sigma_2^2)$, where

$$\begin{aligned}
\sigma_1^2 &= \text{var}\left[\frac{\text{logit}^{-1}\{h(1, X; \beta_0)\}}{\pi_1(1-\pi_1)} - \frac{\text{logit}^{-1}\{h(0, X; \beta_0)\}}{\pi_0(1-\pi_0)}\right], \\
\sigma_2^2 &= \left\{\frac{d_1(\beta_0)}{\pi_1(1-\pi_1)} - \frac{d_0(\beta_0)}{\pi_0(1-\pi_0)}\right\}^{\mathrm{T}} \Sigma(\beta_0) \left\{\frac{d_1(\beta_0)}{\pi_1(1-\pi_1)} - \frac{d_0(\beta_0)}{\pi_0(1-\pi_0)}\right\}, \\
\Sigma(\beta) &= \text{var}\big(\dot{h}(T, X; \beta)[Y - \text{logit}^{-1}\{h(T, X; \beta)\}]\big)^{-1} \\
d_t(\beta) &= E\partial \text{logit}^{-1}\{h(t, X; \beta)\}/\partial\beta \\
&= E\big(\text{logit}^{-1}\{h(t, X; \beta)\}[1 - \text{logit}^{-1}\{h(t, X; \beta)\}]\dot{h}(t, X; \beta)\big),
\end{aligned}$$

and β_0 is the true value of β. Note that $\Sigma(\beta_0)$ is just the asymptotic variance of $\hat{\beta}$, i.e., the inverse of the Fisher information. A consistent variance estimate can be obtained by substituting parameter estimates and sample means and variances in the above display.

Note that the proposed method combines information about two distinct aspects of the population: the covariate distribution and the outcome-covariate relationship, as quantified by β. Both aspects are associated with uncertainty and they contribute to the asymptotic variance of $\hat{\phi}$ in an additive fashion. The uncertainty about the covariate distribution is captured by σ_1^2, which would be the asymptotic variance of the "estimator" should β_0 be known and used in estimating the π_t. The loss of precision due to estimating β_0 is represented by σ_2^2, a quadratic form about the asymptotic variance of $\hat{\beta}$.

4. Numerical Results

4.1. *Simulation Experiments*

A simulation study is conducted to evaluate the finite sample performance of the proposed method as compared to other methods commonly used to estimate a log-OR. In this study, the covariate vector X consists of two binary components (say X_1, X_2) and two continuous ones (X_3, X_4). All four components are independently distributed, with each binary component following the Bernoulli distribution with success probability 0.5, and each continuous component following the standard normal distribution.

Given X, treatment assignment (T) is determined according to the following logistic regression model:

$$\text{logit}(P[T = 1 \mid X]) = a_0 + a_1^{\mathrm{T}} X. \tag{5}$$

In the experiments, a_1 will be set to $(0, 0, 0, 0)$, $(0, 1, 0, 1)$, or $(1, -1, 1, -1)$, with increasing dependence on the covariates. The first value (0 vector) corresponds to a randomized study. For each given value of a_1, a_0 is set to $-E[a_1^{\mathrm{T}}X]$; this implies that $P[T = 1] = 0.5$ because the right-hand side of (5) is symmetrically distributed.

Following assumption (1), the potential outcomes $\{Y(t) : t = 0, 1\}$ are generated as conditionally independent of T given X. Each potential outcome depends on X through a logistic regression model:

$$\mathrm{logit}(P[Y(t) = 1 \mid X]) = b_{t0} + b_{t1}^{\mathrm{T}}X, \quad t = 0, 1.$$

The possible dependence between $Y(0)$ and $Y(1)$ is unidentifiable from the observed data and largely irrelevant in this study. The possible values of b_{01} are $(0, 0, 0, 0)$, $(0, 1, 1, 0)$, and $(1, -1, -1, 1)$, and b_{11} may or may not equal b_{01}. Given the values of b_{01} and b_{11}, the intercepts are determined by $b_{00} = -0.5 - E[b_{01}^{\mathrm{T}}X]$ and $b_{10} = 0.5 - E[b_{11}^{\mathrm{T}}X]$. This does not imply $\pi_0 = P[Y(0) = 1] = \mathrm{logit}^{-1}(-0.5) \approx 0.38$ or $\pi_1 = P[Y(1) = 1] = \mathrm{logit}^{-1}(0.5) \approx 0.62$. Nonetheless, in all scenarios simulated in this study, the marginal probabilities are found, empirically, to stay close to the naive "estimates" given above. Specifically, $0.38 \leq \pi_0 \leq 0.41$, $0.58 \leq \pi_1 \leq 0.62$, and accordingly $0.69 \leq \phi \leq 1.00$.

The actual outcome is of course $Y = Y(T)$. As noted in Sec. 3, the two separate logistic regression models for the potential outcomes $Y(t)$ are jointly equivalent to a single logistic regression model for the actual outcome Y given T and X:

$$\mathrm{logit}(P[Y = 1 \mid T, X]) = \beta_0 + \beta_1 T + \beta_2^{\mathrm{T}}X + \beta_3^{\mathrm{T}}TX, \tag{6}$$

with $\beta_0 = b_{00}$, $\beta_1 = b_{10} - b_{00}$, $\beta_2 = b_{01}$, and $\beta_3 = b_{11} - b_{01}$. The interaction terms are non-null if $b_{01} \neq b_{11}$. A sample consists of $n = 100$ or 500 independent copies of (X, T, Y). In each scenario (combination of sample size and parameter values), 1,000 replicate samples are simulated.

The simulated samples will be used to compare the proposed method with two other methods commonly used to estimate a log-OR. One is based on the sample log-OR derived from a 2×2 table that relates T and Y, ignoring X, with a simple, closed-form standard error (Agresti, 2002). It is well known that the sample log-OR estimates the marginal associational log-OR and may be biased for the log-MCOR due to confounding. This method will be referred to as the direct method. The other method to compare with is based on a logistic regression model, given by (6), for Y given T and X. It returns the regression coefficient for T as the point estimate, with the usual standard error. This method estimates the (constant) conditional OR (and thus is called the conditional method) when there are no interactions between T and X, i.e., when $b_{01} = b_{11}$. When there are interactions between T and X, the regression coefficient for T is hardly interpretable, and this conditional method is not applicable. Both the direct and the conditional methods can be inconsistent, and expressions for their asymptotic bias are given in Appendix B. The proposed method is implemented under model (6), with interactions if necessary. When a sample is analyzed by an applicable method, a point estimate and a standard error are returned, together with an (intended) 95% Wald-type confidence interval.

Table 2 summarizes, in each scenario, the performance of each applicable method in terms of the empirical bias and standard deviation of the point estimator as well as the empirical coverage probability of the associated confidence interval.

Table 2
Empirical comparison of the direct (D), conditional (C), and proposed (P) methods in terms of bias, standard deviation (SD), and coverage probability (CP)

Scenario				Bias × 100			SD × 100			CP (%)		
n	a_1	b_{01}	b_{11}	D	C	P	D	C	P	D	C	P
100	(0, 0, 0, 0)	(0, 0, 0, 0)	(0, 0, 0, 0)	3	8	3	41	44	42	96	95	95
		(0, 1, 1, 0)	(0, 1, 1, 0)	2	29	2	41	51	38	95	93	95
		(1, −1, −1, 1)	(1, −1, −1, 1)	1	40	1	40	58	35	96	90	95
		(0, 1, 1, 0)	(0, 1, 0, 1)	1		1	42		41	94		94
	(0, 1, 0, 1)	(0, 1, 1, 0)	(0, 1, 1, 0)	18	28	1	42	59	44	94	92	93
		(1, −1, −1, 1)	(1, −1, −1, 1)	45	42	2	43	63	40	84	92	95
		(0, 1, 1, 0)	(0, 1, 0, 1)	54		2	43		46	77		95
	(1, −1, 1, −1)	(0, 1, 1, 0)	(0, 1, 1, 0)	44	31	3	44	67	49	82	91	92
		(1, −1, −1, 1)	(1, −1, −1, 1)	−68	43	2	40	71	42	62	90	94
		(0, 1, 1, 0)	(0, 1, 0, 1)	−12		2	45		48	93		93
500	(0, 0, 0, 0)	(0, 0, 0, 0)	(0, 0, 0, 0)	0	1	0	19	19	19	95	95	95
		(0, 1, 1, 0)	(0, 1, 1, 0)	1	23	1	17	20	16	97	83	97
		(1, −1, −1, 1)	(1, −1, −1, 1)	0	33	0	18	22	15	96	69	97
		(0, 1, 1, 0)	(0, 1, 0, 1)	0		0	19		18	94		95
	(0, 1, 0, 1)	(0, 1, 1, 0)	(0, 1, 1, 0)	18	22	0	19	24	19	86	86	95
		(1, −1, −1, 1)	(1, −1, −1, 1)	41	32	−1	18	25	17	41	76	95
		(0, 1, 1, 0)	(0, 1, 0, 1)	52		0	19		20	21		95
	(1, −1, 1, −1)	(0, 1, 1, 0)	(0, 1, 1, 0)	43	23	1	19	25	20	37	86	95
		(1, −1, −1, 1)	(1, −1, −1, 1)	−70	34	1	18	27	17	4	78	94
		(0, 1, 1, 0)	(0, 1, 0, 1)	−14		0	18		20	88		96

In Table 2, the direct method is abbreviated as D, the conditional method as C, and the proposed method as P. Let us fix $n = 100$ for the moment and begin with the very special case where a_1, b_{01}, and b_{11} are all zero. With $a_1 = 0$, treatment assignment is independent of covariates, as in a randomized study, so there is no confounding for the MCOR and the direct method is valid. With $b_{01} = b_{11} = 0$, the potential outcomes are independent of covariates, so the causal ORs are collapsible and the conditional method is valid. Indeed, Table 2 shows that in this situation, all three methods have relatively small biases (for the sample size) and good coverage properties. When b_{01} and b_{11} are non zero, the causal ORs become non collapsible, and the conditional method is clearly biased without interactions and even inapplicable with interactions. On the other hand, the direct method remains valid as long as $a_1 = 0$. It should be noted that, while both the direct and the proposed methods are valid in these few scenarios, the latter method appears slightly more efficient. This possible efficiency gain is not surprising because, unlike the direct method, the proposed method makes use of covariate information. When $a_1 \neq 0$, the direct method becomes biased, too, and the proposed method is the only one that is nearly unbiased. Similar phenomena are observed in larger samples ($n = 500$), except that the valid methods become less biased.

It is mostly for convenience that the log-MCOR, as opposed to the original MCOR, is considered the estimand in this empirical comparison. With minor modifications, all three methods can be used to estimate the original MCOR. In the same simulation experiments of Table 2, the MCOR appears more prone to bias than the log-MCOR when the sample size is small ($n = 100$), after adjusting

for the magnitude of the estimand. However, the difference appears to diminish with increasing sample size and is almost negligible at $n = 500$. In terms of the comparison of the three methods, the remarks in the above paragraph remain valid for estimating the MCOR, qualitatively at $n = 100$ and even quantitatively at $n = 500$.

4.2. *Example*

The methods will now be applied to a real example from Kleinbaum and Klein (2002). The data are from a cohort study in which 609 white males in Evans County, Georgia, were followed for 9 years. The outcome of interest is coronary heart disease status (CHD = 1 if present; 0 if not), and the exposure variable is a dichotomized version of catecholamine level (CAT = 1 if high; 0 if low). The dataset also contains information about a collection of baseline characteristics including AGE, cholesterol level (CHL), smoking status (SMK), electrocardiogram abnormality status (ECG), and hypertension status (HPT). AGE and CHL are continuous, while SMK, ECG, and HPT are binary. Kleinbaum and Klein (2002) described and analyzed this dataset extensively in their discussion of logistic regression.

Without adjusting for baseline characteristics, the sample log-OR for developing CHD in people with high versus low CAT is easily calculated to be 1.05 with a standard error of 0.27. The associated 95% confidence interval is $(0.52, 1.58)$. These statistics provide information about the direct association between CAT and CHD and may not have a causal interpretation due to possible confounding. Kleinbaum and Klein (2002) carefully developed a logistic regression model to assess the impact of CAT on CHD while adjusting for potential confounders. The model includes as covariates CAT, CHL, SMK, ECG, and HPT, with two interaction terms: CAT $\times$ CHL and CAT $\times$ HPT. Because of the interaction terms, the conditional associational OR is not assumed constant. Based on the fitted model, the conditional associational log-OR is estimated to be $-12.69 + 0.07 \times \text{CHL} - 2.33 \times \text{HPT}$, which also estimates the conditional causal log-OR under the assumption of no unmeasured confounders. The proposed method is applied under the same logistic regression model to yield a point estimate of 1.30 with a standard error of 0.30. The associated 95% confidence interval is $(0.71, 1.89)$. These results appear to suggest a stronger marginal causal relationship than reflected in the observed association between CAT and CHD, again under the assumption of no unmeasured confounders.

5. Discussion

Although the issues of confounding and collapsibility have long been recognized, it seems worthwhile to highlight their ramifications in the context of ORs. Table 1 clarifies the subtle differences between several possible ways to define an OR. These differences should be kept in mind when choosing an estimation method and when interpreting an estimated OR. It is somewhat surprising that estimation of an MCOR subject to confounding has not been discussed as much as the other entries of Table 1. This article proposes a method based on a logistic regression model relating the observed outcome to the received treatment and potential confounders. Simulation results show that the proposed method performs reasonably well in moderate-sized samples and may even offer an efficiency gain over the direct method based on the sample OR in the absence of confounding.

Obviously, there are other possible approaches to this estimation problem. For example, it is possible to estimate the π_t using an inverse probability weighting approach with weights obtained from a propensity score model. It is also possible to construct doubly robust estimators that are consistent and asymptotically normal if either the outcome regression model or the propensity score model is correctly specified (e.g., Bang and Robins, 2005). However, there appear to be numerical difficulties with methods that involve weighting observations by inverses of possibly near-zero probabilties.

Besides the MCOR, the marginal causal effect of a binary exposure on a binary outcome could also be described in terms of the relative risk (π_1/π_0) or a simple difference ($\pi_1 - \pi_0$). The latter quantities can be easily estimated under the logistic regression framework of this article, by substituting the estimates $\hat{\pi}_0$ and $\hat{\pi}_1$ from Sec. 3. It should be noted, however, that the logit link assumed in this article may become less attractive when estimating the relative risk or the difference. It may appear more natural to use the log link for the relative risk, or the identity link for the difference. Choosing the right link function for the effect measure of interest also helps to alleviate concerns about potential non collapsibility (Gail et al., 1984).

A prospective design has been assumed in this article. Under a retrospective design, all of β cannot be estimated. Specifically, if $h(T, X; \beta)$ in model (4) is linear in β, then the intercept cannot be identified from the retrospective data while the other regression coefficients can. It appears difficult to estimate the MCOR in this situation. However, it may be possible to identify and estimate the MCOR with additional information, say the prevalence $P[Y = 1]$. Further research is warranted.

Appendix A: Asymptotic Theory for $\hat{\phi}$

Standard regularity conditions in the M-estimation theory (e.g., van der Vaart, 1998, Ch. 5) are assumed. These include parameter identifiability in the logistic regression model, smoothness of the model in parameters, positivity of the Fisher information, existence of integrable envelopes that permit use of the dominated convergence theorem, and certain Donsker properties that help deal with random functions. Techniques for verifying the Donsker property can be found in van der Vaart and Wellner (1996). Let $\mathbb{P}_0$ denote the true distribution of (T, X, Y) and $\mathbb{P}_n$ the empirical distribution of the (T_i, X_i, Y_i), $i = 1, \ldots, n$. Write $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - \mathbb{P}_0)$ for the empirical process. We shall use operator notation for integrals, writing, for instance,

$$\hat{\pi}_t = \mathbb{P}_n \text{logit}^{-1}\{h(t, X; \hat{\beta})\}, \quad t = 0, 1.$$

It is well established that

$$\sqrt{n}(\hat{\beta} - \beta_0) = -\Sigma(\beta_0)\mathbb{G}_n\big(\dot{h}(T, X; \beta_0)[Y - \text{logit}^{-1}\{h(T, X; \beta_0)\}]\big) + o_p(1). \quad (7)$$

Let us write, for $t = 0, 1$,

$$\begin{aligned}\sqrt{n}(\hat{\pi}_t - \pi_t) &= \sqrt{n}\big[\mathbb{P}_n \text{logit}^{-1}\{h(t, X; \widehat{\beta})\} - \mathbb{P}_0 \text{logit}^{-1}\{h(t, X; \beta_0)\}\big] \\ &= \mathbb{G}_n \text{logit}^{-1}\{h(t, X; \widehat{\beta})\} \\ &\quad + \sqrt{n}\mathbb{P}_0\big[\text{logit}^{-1}\{h(t, X; \widehat{\beta})\} - \text{logit}^{-1}\{h(t, X; \beta_0)\}\big]. \quad (8)\end{aligned}$$

By the dominated convergence theorem, the map

$$\beta \mapsto \text{logit}^{-1}\{h(t, X; \beta)\} \in L_2(\mathbb{P}_0)$$

is continuous at β_0. Further, by the continuous mapping theorem,

$$\|\text{logit}^{-1}\{h(t, X; \widehat{\beta})\} - \text{logit}^{-1}\{h(t, X; \beta_0)\}\|_{2,\mathbb{P}_0} = o_p(1),$$

where the L_2 norm is evaluated under the true distribution of (T, X, Y) with $\widehat{\beta}$ regarded as an index. It then follows from theorem 19.24 of van der Vaart (1998) that

$$\mathbb{G}_n\text{logit}^{-1}\{h(t, X; \widehat{\beta})\} = \mathbb{G}_n\text{logit}^{-1}\{h(t, X; \beta_0)\} + o_p(1). \tag{9}$$

Again by the dominated convergence theorem, the map

$$\beta \mapsto \mathbb{P}_0\text{logit}^{-1}\{h(t, X; \beta)\}$$

is differentiable at β_0 with derivative $d_t(\beta_0)$, $t = 0, 1$. Hence,

$$\sqrt{n}\,\mathbb{P}_0\big[\text{logit}^{-1}\big\{h(t, X; \widehat{\beta})\} - \text{logit}^{-1}\{h(t, X; \beta_0)\big\}\big] = d_t(\beta_0)^{\mathrm{T}}\sqrt{n}(\widehat{\beta} - \beta_0) + o_p(1), \tag{10}$$

by the delta method. In view of (9), (10), and (7), Eq. (8) becomes

$$\begin{aligned}\sqrt{n}(\widehat{\pi}_t - \pi_t) = \mathbb{G}_n\big(&\text{logit}^{-1}\{h(t, X; \beta_0)\} \\ &- d_t(\beta_0)^{\mathrm{T}}\Sigma(\beta_0)\dot{h}(T, X; \beta_0)[Y - \text{logit}^{-1}\{h(T, X; \beta_0)\}]\big) + o_p(1).\end{aligned}$$

Apply the delta method once again to obtain

$$\begin{aligned}\sqrt{n}(\widehat{\phi} - \phi) &= \frac{\sqrt{n}(\widehat{\pi}_1 - \pi_1)}{\pi_1(1 - \pi_1)} - \frac{\sqrt{n}(\widehat{\pi}_0 - \pi_0)}{\pi_0(1 - \pi_0)} + o_p(1) \\ &= \mathbb{G}_n\{u_1(X) - u_2(T, X, Y)\} + o_p(1),\end{aligned}$$

where

$$u_1(X) = \frac{\text{logit}^{-1}\{h(1, X; \beta_0)\}}{\pi_1(1 - \pi_1)} - \frac{\text{logit}^{-1}\{h(0, X; \beta_0)\}}{\pi_0(1 - \pi_0)},$$

$$u_2(T, X, Y) = \left\{\frac{d_1(\beta_0)}{\pi_1(1 - \pi_1)} - \frac{d_0(\beta_0)}{\pi_0(1 - \pi_0)}\right\}^{\mathrm{T}}\Sigma(\beta_0)\dot{h}(T, X; \beta_0)[Y - \text{logit}^{-1}\{h(T, X; \beta_0)\}].$$

Note that $u_1(X)$ and $u_2(T, X, Y)$ are uncorrelated because $E[u_2(T, X, Y) \mid T, X] \equiv 0$. It follows that

$$\sqrt{n}(\widehat{\phi} - \phi) \to \text{N}(0, \sigma^2)$$

with

$$\sigma^2 = \text{var}\{u_1(X)\} + \text{var}\{u_2(T, X, Y)\} = \sigma_1^2 + \sigma_2^2.$$

Appendix B: Asymptotic Bias of the Direct and the Conditional Methods

In the setting of Sec. 4.1, the log-MCOR is given by $\phi = \text{logit}(\pi_1) - \text{logit}(\pi_0)$, where

$$\pi_t = P[Y(t) = 1] = E\{P[Y(t) = 1 \mid X]\} = E\{\text{logit}^{-1}(b_{t0} + b_{t1}^{\mathrm{T}} X)\}, \quad t = 0, 1.$$

If $b_{01} = b_{11}$, then the conditional method estimates $b_{10} - b_{00}$, and its asymptotic bias is just $b_{10} - b_{00} - \phi$ with ϕ given above. If $b_{01} \neq b_{11}$, then there are treatment-covariate interactions and the conditional method is not applicable. The direct method, on the other hand, estimates $\phi^* = \text{logit}(\pi_1^*) - \text{logit}(\pi_0^*)$, where

$$\begin{aligned}
\pi_t^* &= P[Y = 1 \mid T = t] = P[Y(t) = 1 \mid T = t] = E\{P[Y(t) = 1 \mid X, T = t] \mid T = t\} \\
&= E\{P[Y(t) = 1 \mid X] \mid T = t\} \quad \text{by assumption (1)} \\
&= E\{\text{logit}^{-1}(b_{t0} + b_{t1}^{\mathrm{T}} X) \mid T = t\}, \quad t = 0, 1.
\end{aligned}$$

Let us write f for the density of X with respect to some measure μ, and f_t for the conditional density of X given $T = t$. Then, by Bayes' rule,

$$\begin{aligned}
f_t(x) &= \frac{f(x)P[T = t \mid X = x]}{\int f(z)P[T = t \mid X = z]d\mu(z)} \\
&= \frac{f(x)\{\text{logit}^{-1}(a_0 + a_1^{\mathrm{T}} x)\}^t \{1 - \text{logit}^{-1}(a_0 + a_1^{\mathrm{T}} x)\}^{1-t}}{\int f(z)\{\text{logit}^{-1}(a_0 + a_1^{\mathrm{T}} z)\}^t \{1 - \text{logit}^{-1}(a_0 + a_1^{\mathrm{T}} z)\}^{1-t} d\mu(z)}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\pi_t^* &= \int \text{logit}^{-1}(b_{t0} + b_{t1}^{\mathrm{T}} x) f_t(x) d\mu(x) \\
&= \frac{\int f(x)\{\text{logit}^{-1}(a_0 + a_1^{\mathrm{T}} x)\}^t \{1 - \text{logit}^{-1}(a_0 + a_1^{\mathrm{T}} x)\}^{1-t} \text{logit}^{-1}(b_{t0} + b_{t1}^{\mathrm{T}} x) d\mu(x)}{\int f(z)\{\text{logit}^{-1}(a_0 + a_1^{\mathrm{T}} z)\}^t \{1 - \text{logit}^{-1}(a_0 + a_1^{\mathrm{T}} z)\}^{1-t} d\mu(z)}.
\end{aligned}$$

The asymptotic bias of the direct method is given by $\phi^* - \phi$.

Acknowledgments

The views expressed in this article do not necessarily represent those of the U.S. Food and Drug Administration.

References

Agresti, A. (2002). *Categorical Data Analysis*. 2nd ed. New York: Wiley.

Austin, P. C. (2007). The performance of different propensity score methods for estimating marginal odds ratios. *Statist. Med.* 26:3078–3094.

Austin, P. C., Grootendorst, P., Normand, S.-L. T., Anderson, G. M. (2007). Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: A Monte Carlo study. *Statist. Med.* 26:754–768.

Bang, H., Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* 61:962–972.

D'Agostino, R. B. Jr. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statist. Med.* 17:2265–2281.

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

Greenland, S., Robins, J. M., Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statist. Sci.* 14(1):29–46.

Kleinbaum, D. G., Klein, M. (2002). *Logistic Regression*. New York: Springer-Verlag.

Little, R. J. A., Rubin D. B. (2002). *Statistical Analysis with Missing Data*. 2nd ed. New York: Wiley.

Lunceford, J. K., Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statist. Med.* 23:2937–2960.

Robins, J. M., Ritov, Y. (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statist. Med.* 16:285–319.

Rosenbaum, P. R., Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55.

Rosenbaum, P. R., Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *JASA* 79:516–524.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* 63:581–592.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

van der Vaart, A. W., Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. New York: Springer-Verlag.