

Estimation of Risk Ratios in Cohort Studies  
With Common Outcomes: A Simple and  
Efficient Two-stage Approach

Eric J. Tchetgen\*

\*Harvard University, [etchetge@hsph.harvard.edu](mailto:etchetge@hsph.harvard.edu)

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/harvardbiostat/paper136>

Copyright ©2011 by the author.

# Estimation of Risk Ratios in Cohort Studies With Common Outcomes: A Simple and Efficient Two-stage Approach

Eric J. Tchetgen Tchetgen

Departments of Epidemiology and Biostatistics,  
Harvard University

Correspondence: Eric J. Tchetgen Tchetgen, Department of Epidemiology, Harvard School of Public Health 677 Huntington Avenue, Boston, MA 02115.

## Abstract

The risk ratio is perhaps the effect measure most commonly assessed in epidemiologic studies with a binary outcome. In this paper, the author presents a simple and efficient two-stage approach to estimate risk ratios directly, which does not directly rely for consistency on an estimate of the baseline risk. This latter property is a key advantage of the approach over existing methods, because, unlike these other methods, the proposed approach obviates the need to restrict the predicted risk probabilities to fall below one, in order to recover efficient inferences about risk ratios. An additional appeal of the approach is that it is easy to implement. Finally, when the primary interest is in the effect of a specific binary exposure, a simple doubly robust closed-form estimator is derived, for the multiplicative effect of the exposure. Specifically, we show how one can adjust for confounding by incorporating a working regression model for the propensity score so that correct inferences about the multiplicative effect of the exposure are recovered if either this model is correct or a working model for the association between confounders and outcome risk is correct, but both do not necessarily hold.

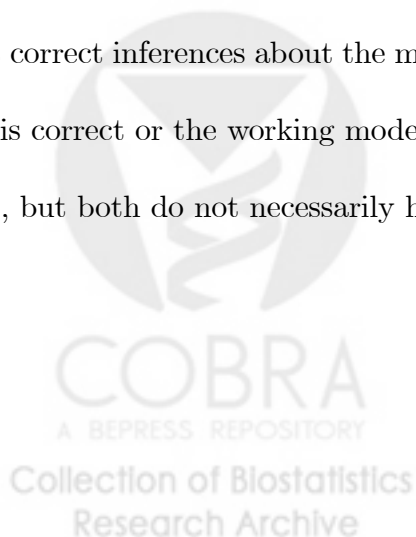
KEY WORDS: Risk ratio, prevalence ratio, semiparametric efficient, doubly robust

# 1 Introduction

An objective of many epidemiologic studies is to evaluate the multiplicative association between a vector of risk factors and a binary outcome. When the outcome is rare within all levels of the covariates, logistic regression is well known to deliver valid, albeit approximate, inferences about risk ratios whether in a cohort or in a case-control study. When, as often the case in cohort studies, the outcome is not rare within all levels of covariates, logistic regression overstates the relative risk association and should not be used to approximate the latter. Instead, a variety of techniques have been proposed in recent years to recover estimates of risk ratios for a common outcome (Wacholder, 1986, Lee, 1994, Skov et al, 1998, Greenland, 2004, Zou, 2004, Spiegelman and Hertzmark, 2005, Chu and Cole, 2010). A basic requirement shared by previous methods, with the exception of the method proposed by Breslow (1974) and subsequently by Lee (2004) is that the log-baseline risk, i.e. the regression intercept, must be estimated along with regression coefficients, in order to obtain a consistent estimate of regression coefficients. Unfortunately, this task is often not easily achieved if one wishes to respect the essential model restriction that all predicted probabilities in the sample should not exceed one; often resulting in lack of convergence of estimation procedures. The suboptimal performance of such methods are well documented in the literature (Deddens et al, 2003, Petersen and Deddens, 2006, Tian and Liu, 2006, Chu and Cole, 2010). Recently, such concerns prompted Chu and Cole to develop a Bayesian approach that appropriately incorporates this additional modeling restriction (Chu and Cole, 2010). Their approach which relies on Markov Chain Monte Carlo simulations provides a promising Bayesian solution when risk prediction is of primary interest, but a satisfactory frequentist solution is still lacking even in settings where risk ratios are the primary target of inference.

In this paper, the author presents a simple approach to estimate risk ratios directly, that does

not directly rely for consistency on obtaining an estimate of the baseline risk. In this respect, the approach is similar to that of Breslow (1974) and Lee (2004); but whereas their method is inefficient, here a two-stage approach is described that delivers efficient estimates of risk ratios. The first stage of the method does not require an estimate of the baseline risk, while the second stage recovers information not used in the first stage by incorporating a weight which does depend on the individual predicted risk, and therefore on the individual baseline risk. However, because the weights are not essential for consistency, a simple pluggin estimate of the baseline risk may be used without altering the large sample behavior, more precisely, without altering the large sample efficiency of the estimated regression coefficients. This property holds even though the pluggin estimate is generally inefficient for the baseline risk and may result in a predicted risk outside of the unit range. An important advantage of the approach is that it is easy to implement. An alternative approach is described, which guarantees that the estimated predicted risk used for the weight remains bounded between zero and one. Finally, when the primary interest is in the effect of a specific binary exposure, we describe a simple closed-form estimator, of the multiplicative effect of the exposure that is doubly robust. Specifically, we show how to incorporate a working regression model for the probability of being exposed given confounders, i.e. the propensity score, so that correct inferences about the multiplicative effect of the exposure are recovered if either this model is correct or the working model for the association between confounders and disease risk is correct, but both do not necessarily hold.



## 2 PROPOSED METHODS

### 2.1 A Simple Inefficient Initial Estimator

To motivate the approach, consider the simple case where  $X_i$  ( $i = 1, 2, \dots, n$ ) is a binary exposure with a value of 1 if exposed and 0 if unexposed. Let  $Y_i$  ( $i = 1, \dots, n$ ) denote the binary response, which is randomly sampled from a log-binomial model with

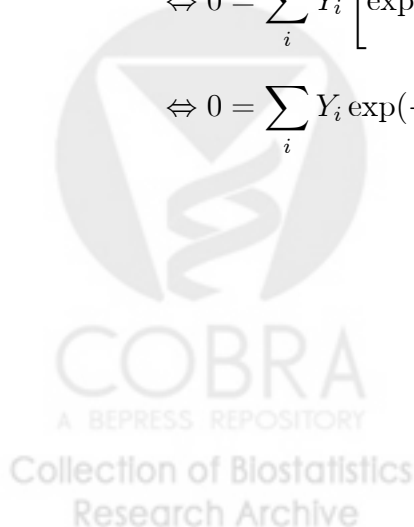
$$\log \Pr(Y_i = 1|X_i) = \log[p_i] = \alpha_0 + \beta_0 X_i \quad (1)$$

Then, a standard application of maximum likelihood theory delivers the estimator

$$\exp(\hat{\beta}_{MLE}) = \frac{\sum_i Y_i X_i}{\sum_i X_i} \times \frac{\sum_i (1 - X_i)}{\sum_i (1 - X_i) Y_i}$$

Now, we note that this equation is equivalent to:

$$\begin{aligned} 0 &= \exp(-\hat{\beta}_{MLE}) \sum_i Y_i X_i \sum_i (1 - X_i) - \sum_i (1 - X_i) Y_i \sum_i X_i \\ \Leftrightarrow 0 &= \sum_i Y_i \left[ \exp(-\hat{\beta}_{MLE}) X_i \sum_i (1 - X_i) - (1 - X_i) \sum_i X_i \right] \\ \Leftrightarrow 0 &= \sum_i Y_i \exp(-\hat{\beta}_{MLE} X_i) \left[ X_i n - \sum_i X_i \right] \end{aligned}$$



which states that  $\hat{\beta}_{MLE}$  solves the equation

$$\begin{aligned}
 0 &= \sum_i Y_i \exp(-\hat{\beta}_{MLE} X_i) (X_i - \bar{X}) \\
 \Leftrightarrow 0 &= \sum_i Y_i \exp(-\hat{\beta}_{MLE} (X_i - \bar{X})) (X_i - \bar{X}) \\
 \Leftrightarrow 0 &= \sum_{i:Y_i=1} \left\{ Z_i - \exp(\hat{\beta}_{MLE} W_i) \right\} W_i
 \end{aligned} \tag{2}$$

where  $\bar{X}$  is the sample average of  $X$ ,  $W_i = -(X_i - \bar{X})$ , and  $Z_i = 0$  for all  $i$ . The main appeal of the representation given by equation (2) in the above display is two-fold:

- (i) It is completely free of the intercept, and therefore does not require an actual estimate of the predicted probabilities.
- (ii) It is exactly of the form of the score equation for  $\beta$ , under the artificial case-only model in which the pseudo-outcome  $Z_i$  is assumed to follow a Poisson distribution with mean given by the intercept-free multiplicative model  $\exp(\beta W_i)$ ,  $i = 1, \dots, n$ , in cases only.

Thus, Equation (2) provides an equivalent representation of the maximum likelihood estimator in the simple setting of a saturated multiplicative model with a binary exposure; however, this representation is of no particular use in this latter setting because the maximum likelihood estimator is easy to compute. But, as we show below, the alternative representation is useful for estimation in settings where it may be considerably more difficult to compute the maximum likelihood estimator. Specifically, now suppose that  $X_i$  and thus  $W_i$ , are vector valued possibly with several continuous components and one aims to make inferences about  $\beta_0$  in the multiplicative model

$$\frac{\Pr(Y_i = 1|X_i)}{\Pr(Y_i = 0|X_i)} = \exp \{ \beta_0^T X_i \} \tag{3}$$

Then, one may generalize equation (2), and define an estimator  $\hat{\beta}$  as the solution to the equation:

$$\begin{aligned} 0 &= \sum_{i:Y_i=1} U_i(\hat{\beta}) \\ &= \sum_{i:Y_i=1} \left\{ Z_i - \exp(\hat{\beta}^T W_i) \right\} W_i \end{aligned}$$

In the appendix, we show that  $\hat{\beta}$  is consistent for  $\beta_0$  and we establish its large sample behavior.

*Result 1: Under assumption (1),  $n^{1/2}(\hat{\beta} - \beta_0)$  is approximately normal with mean zero and variance  $\Sigma_\beta$  provided in the appendix. We also show that the standard sandwich estimator*

$$\hat{\Sigma}_\beta = \left[ \sum_{i:Y_i=1} \frac{\partial U_i(\beta)}{\partial \beta} \Big|_{\hat{\beta}} \right]^{-1} n \sum_{i:Y_i=1} U_i(\hat{\beta})^2 \left[ \sum_{i:Y_i=1} \frac{\partial U_i(\beta)}{\partial \beta} \Big|_{\hat{\beta}} \right]^{-1}$$

*is a conservative estimator of  $\Sigma_\beta$ .*

The estimator  $\hat{\beta}$  is particularly useful for routine application in epidemiologic practice, because properties (i) and (ii) continue to apply even though model (3) is no longer saturated, and therefore  $\hat{\beta}$  does not generally inherit the efficiency properties of a maximum likelihood estimator. The efficiency loss (relative to a maximum likelihood estimator) can be particularly severe when the regression model is not saturated, and when as we assume throughout, the outcome is not rare. The loss of efficiency should decrease the more flexible or richly parametrize the model is allowed to remain, and should be almost null for nearly saturated models. Despite this limitation, the approach has some advantages in that by (i) it does not require an estimate of the intercept and therefore will generally not suffer from the same computational challenges as methods that rely on an estimate of the intercept. For inference using  $\hat{\beta}$ , valid confidence intervals, for say the first component  $\beta_0^{(1)}$  of  $\beta_0$ , can be obtained by the method of Wald :  $\hat{\beta}^{(1)} \pm 1.96 \sqrt{\hat{\Sigma}_\beta^{11}/n}$ , where  $\hat{\Sigma}_\beta^{11}/n$

is the estimate of the variance of  $\widehat{\beta}^{(1)}$ . This approach is convenient, as (ii) outlines how to obtain  $\widehat{\beta}$  using standard statistical software such as GENMOD; which also provides the empirical/sandwich variance estimator  $\widehat{\Sigma}_\beta$  upon request, i.e by specifying the REPEATED statement.

We performed a simulation study to illustrate the performance of the method. For this we generated 1000 samples each of size  $n = 1000$ , under the following model  $X^{(2)}$  is Bernoulli (0.7),  $X^{(3)}$ ,  $X^{(4)}$  are both uniform(0,1);  $X^{(1)}$  is Bernoulli( $(1 + \exp(-[0.5, -0.5, 0.5, -0.9, 0.9] \times Q))$ ) where  $Q = [1, X^{(2)} \times X^{(3)}, X^{(3)} \times X^{(4)}, X^{(2)} \times X^{(4)^2}]$ ;  $Y$  is then generated under a Bernoulli model with event probability  $\exp([-1.4, 0.3, -0.2, -0.2, 0.3] \times [1, X^T]^T)$ , thus  $[\alpha_0, \beta_0^T] = [-1.4, 0.3, -0.2, -0.2, 0.3]$  which roughly corresponds to a marginal risk  $\Pr(Y = 1) \approx 0.278$ . We then obtained estimates using the method described in this section which we summarize in Table 1 in rows labelled "Correct Model". The simulation study indicates that the point estimate  $\widehat{\beta}$  performs well and has small bias. The simulation further shows that the simple sandwich estimator  $\widehat{\Sigma}_\beta$  can be quite conservative as it produces estimates that can be much larger than the Monte Carlo variance. Instead of using  $\widehat{\Sigma}_\beta$ , alternative inferences can also be obtained by using an empirical version of  $\Sigma_\beta$ , which we denote  $\widetilde{\Sigma}_\beta$  and is given by

$$\begin{aligned} \widetilde{\Sigma}_\beta = & \widehat{\Sigma}_\beta - \left[ \sum_i Y_i \{ \exp(-\beta_0^T (X_i - \bar{X})) \} \right]^2 \\ & \times \left[ \sum_{i:Y_i=1} \frac{\partial U_i(\beta)}{\partial \beta} \Big|_{\widehat{\beta}} \right]^{-1} n \sum_{i:Y_i=1} (X_i - \bar{X})^{\otimes 2} \left[ \sum_{i:Y_i=1} \frac{\partial U_i(\beta)}{\partial \beta} \Big|_{\widehat{\beta}} \right]^{-1} \end{aligned}$$

as derived in the appendix. However, this more precise estimator may be less convenient as it requires additional, though fairly straightforward programming. The simulation study indicates that  $\widetilde{\Sigma}_\beta$  outperforms  $\widehat{\Sigma}_\beta$  and performs well.



## 2.2 An Efficient Estimator

To address concerns about lack of efficiency, suppose that we have obtained  $\widehat{\beta}$  in a first stage. One can then update  $\widehat{\beta}$  in a single step, to obtain an efficient estimator of  $\beta_0$ . Let

$$\widehat{T}_i(w) = \left\{ w_i - \frac{\sum_i w_i \exp(\widehat{\beta}^T X_i)}{\sum_i \exp(\widehat{\beta}^T X_i)} \right\}$$

where  $w_i$  is a vector, of the same dimension as  $X_i$ , of user-specified functions of  $X_i$ . For any choice of  $w_i$ , let

$$\widehat{\beta}(w) = \widehat{\beta} + \left[ \sum_i Y_i \widehat{T}_i(w) X_i^T \right]^{-1} \times \left[ \sum_i Y_i \widehat{T}_i(w) \right]$$

define a new so-called one-step-update estimator. The class of one-step-update estimators is very rich and includes several well-known estimators. In fact, for any estimator  $\bar{\beta}$  of  $\beta_0$  that is regular and asymptotically linear, we show in the appendix using results due to Bickel et al (1993), that there exist a corresponding weight function  $w_i$  such that

$$\sqrt{n} \left\{ \widehat{\beta}(w) - \bar{\beta} \right\} = o_p(1)$$

In other words, the two estimators share a common large sample distribution and are therefore asymptotically equivalent. For instance, one can easily verify that the particular choice  $w_i = \exp(-\widehat{\beta}^T X_i)(X_i - \bar{X})$  recovers  $\widehat{\beta}$  exactly. Whereas,  $w_i = X_i$  produces an estimator that is asymptotically equivalent to the Breslow-Lee estimator. Neither of these estimators is generally efficient. In the appendix, we show that  $\widehat{\beta}(w_{opt})$  is efficient, where

$$w_{opt,i} = \{1 - \widehat{p}_i\}^{-1} \times \left[ X_i - \frac{\sum_i X_i \{1 - \widehat{p}_i\}^{-1} \widehat{p}_i}{\sum_i \{1 - \widehat{p}_i\}^{-1} \widehat{p}_i} \right]$$

with

$$\hat{p}_i = \exp(\hat{\beta}^T X_i) \sum_{i'} Y_{i'} \exp(-\hat{\beta}^T X_{i'}) / n$$

an estimator of the predicted risk for person  $i$ ; where

$$\exp(\hat{\alpha}) = \sum_{i'} Y_{i'} \exp(-\hat{\beta}^T X_{i'}) / n$$

is a pluggin estimator of the baseline risk  $\Pr(Y = 1|X = 0) = \exp(\alpha_0)$ . Specifically, we establish that

$$\hat{\beta}(w_{opt}) = \hat{\beta} + \left[ \sum_i Y_i w_{opt,i} X_i^T \right]^{-1} \times \left[ \sum_i Y_i w_{opt,i} \right]$$

since  $\sum_i w_{opt,i} \exp(\hat{\beta}^T X_i) = 0$ . In fact, we prove the following result:

*Result 2: Under assumption (1),  $n^{-1/2} (\hat{\beta}_{eff} - \beta_0)$  is approximately normal with mean zero and variance  $\Sigma_{\beta}^{eff}$ . Furthermore, the estimator  $\hat{\Sigma}_{\beta}^{eff}$  converges (in probability) to  $\Sigma_{\beta}^{eff}$  where*

$$\hat{\Sigma}_{\beta}^{eff} = n \left[ \sum_i \{1 - \hat{p}_i\} \hat{p}_i w_{opt,i} w_{opt,i}^T \right]^{-1}$$

Finally,  $\hat{\beta}_{eff}$  achieves the semiparametric efficiency bound for the model given by (1).

As before,  $\hat{\Sigma}_{\beta}^{eff}$  can be used to construct Wald-type confidence intervals. The simulation results in table 1 confirm that, as theory predicts  $\hat{\beta}_{eff}$  significantly outperforms  $\hat{\beta}$  in terms of efficiency. We emphasize that the estimated individual risk  $\hat{p}_i, i = 1, \dots, n$ , is solely used for the purpose of enhancing efficiency through the weights  $w_{opt,i}$ . Result 2 confirms that the baseline log-risk  $\alpha_0$  may be inefficiently estimated by the simple pluggin estimator  $\hat{\alpha}$ , without affecting the efficiency of  $\hat{\beta}_{eff}$ . However, although  $\hat{\alpha}$  is consistent and asymptotically linear,  $\hat{p}_i$  may be greater than one for some observations in the sample. Naturally, one may wish to impose that the estimated risk used

to compute the optimal weight be a genuine probability; in the next section, we describe a slight modification of the proposed approach that achieves this goal.

### 3 Additional results and an application

#### 3.1 An alternative efficient estimator

While not strictly required by the two stage approach, the following modification guarantees that individuals' estimated risk used to compute the weights  $w_{opt,i}$  fall within the unit interval. To develop the approach, we observe that  $p_i$  is equivalently written:

$$\text{logit}p_i = \xi(\beta_0^T X_i)$$

where

$$\xi(\cdot) = \log \{ \exp(\alpha_0 + \cdot) / (1 - \exp(\alpha_0 + \cdot)) \}.$$

Given the first stage estimate  $M_i = \hat{\beta}^T X_i$  of  $\beta_0^T X_i$ , we propose to ignore knowledge about the precise functional form of  $\xi(\cdot)$ , and to estimate  $\xi(\cdot)$  by fitting a nonparametric logistic regression of  $Y_i$  on the scalar variable  $M_i, i = 1, \dots, n$ . Let  $\hat{\xi}_i = \hat{\xi}(M_i)$  denotes such an estimator of  $\xi(\beta_0^T X_i)$ ; then clearly

$$0 < \tilde{p}_i = \left\{ 1 + \exp \left( - \left( \hat{\xi}_i \right) \right) \right\}^{-1} < 1, i = 1, \dots, n$$

that is  $\tilde{p}_i$  is guaranteed to fall within the unit interval. There currently exist a vast literature on nonparametric techniques that may be used to obtain  $\hat{\xi}(\cdot)$ , including polynomials series, local polynomial smoothing, trigonometric series, wavelet regression, spline regression or kernel smoothing; a textbook treatment of these various methods may be found in Wasserman (2003) and Hastie

et al (2008). Here, we briefly illustrate polynomial series regression. Let  $\phi_k(M_i) = M_i^k, k = 0, \dots, K$ . Then, for fixed  $K$ , let  $\tilde{p}_i$  denote the predicted probabilities obtained by standard logistic regression of  $Y_i$  on  $\{\phi_k(M_i); k \leq K\}$  using data  $\{(M_i, Y_i) : i = 1, \dots, n\}$ . A result due to Hirano et al (2003) implies that, since  $\xi(\cdot)$  has at least four bounded derivatives, setting  $K = Cn^{1/6}$  for some constant  $C$  is sufficient for the resulting estimator  $\tilde{p}_i$  to converge to  $p_i$  at rates no slower than  $n^{1/4}$ ; and the resulting estimator  $\tilde{\beta}_{eff}$  of  $\beta_0$  is semiparametric efficient.

### 3.2 A data illustration

We consider a data set involving 172 diabetic patients presented by Lachin (14, p. 261) and also analyzed by Zou (2003). This is a subset of a large clinical trial known as the Diabetes Control and Complications Trial (The Diabetes Control and Complications Trial Research Group, 1993), where it is of interest to determine the relative risk of standard therapy versus intensive treatments in terms of the prevalence of microalbuminuria at 6 years of follow-up. For estimation, we adjust for the following covariates: the percentage of total hemoglobin that has become glycosylated at baseline, the prior duration of diabetes in months, the level of systolic blood pressure (mmHg), and gender (female) (1 if female, 0 if male). Applying the single stage approach results in an estimated risk of microalbuminuria that is 2.5 times higher in the control group than in the treatment group ( $\hat{\beta} = -0.92, s.e. = 0.37$ ). The efficient two-stage approach delivers a more precise estimated risk ratio, with the risk in the control group that is 5.4 times higher than in the treatment group ( $\hat{\beta}_{eff} = -1.69, s.e. = 0.28$ ) using the simple pluggin approach for estimating individuals' predicted risk, and an estimated risk that is 3.2 times higher in the control group ( $\tilde{\beta}_{eff} = -1.18, s.e. = 0.25$ ) using the approach described in Section 5.4. It is Interest to compare these point estimates to those reported by Zou (2003) who estimated that the risk in the control group is 2.9 that in the treatment group ( $\hat{\beta}_{Zou} = -1.08, s.e. = 0.30$ ) using a modified Poisson approach, which closely matched the

estimated risk ratio of 2.85 for the control vs the treatment group ( $\widehat{\beta}_{bin} = -1.04, s.e. = 0.30$ ) he obtained using the log-binomial approach. He further noted that the binomial regression procedure failed to converge until a variety of starting values were provided, when it finally converged with a starting value of  $-1.1$  for the intercept. The two-stage estimator appears to provide more precise inference about the treatment effect than the other methods.

### 3.3 Double robustness

Suppose that, as often the case in epidemiologic studies, we are particularly interested in the effect  $\beta^{(1)}$  of the first component  $X^{(1)}$  of  $X$ , which represents a binary exposure under study, and the remaining sub-vector  $X^{(-1)}$  of  $X$  includes confounding factors with corresponding effect  $\beta^{(-1)}$ , so that  $X = (X^{(1)}, X^{(-1)T})^T$  and  $\beta = (\beta^{(1)}, \beta^{(-1)T})^T$ . Then, strictly speaking  $\beta^{(-1)}$  is a nuisance parameter not of direct interest, and the model

$$\frac{\Pr(Y_i = 1 | X^{(1)} = 0, X^{(-1)})}{\Pr(Y_i = 1 | X_i = 0)} = \exp \left\{ \beta_0^{(-1)T} X_i \right\} \quad (4)$$

is a working model used strictly for the purpose of confounding adjustment. Unless the working model in the display above is saturated, in general one cannot rule out possible model misspecification which in turn can result in biased inferences about the exposure effect, due to inadequate confounding adjustment. Because saturated models will generally be impractical due to data sparseness, we propose to partially alleviate these concerns by modeling the probability of exposure given covariates, i.e. the propensity score, with a working regression model,

$$\text{logit}\{\Pr(X^{(1)} = 1 | X^{(-1)}; \psi_0)\} = \psi_0^T [1, X^{(-1)}]^T \quad (5)$$

Suppose  $\widehat{\psi}$  is the maximum likelihood estimator of  $\psi_0$ , and let  $\widehat{\pi}_i = \Pr(X^{(1)} = 1 | X^{(-1)}; \widehat{\psi})$ . Then  $\widehat{\beta}_{dr}^{(1)}$  is doubly robust, where

$$\widehat{\beta}_{dr}^{(1)} = \log \frac{\sum_i Y_i X_i^{(1)} \exp(-\beta^{(-1)T} X^{(-1)}) \{1 - \widehat{\pi}_i\}}{\sum_i Y_i (1 - X_i^{(1)}) \exp(-\beta^{(-1)T} X^{(-1)}) \widehat{\pi}_i}$$

that is

*Result 3:  $\widehat{\beta}_{dr}^{(1)}$  converges (in probability) to  $\beta_0^{(1)}$  if either model (4) holds, or model (5) holds, but not necessarily both hold.*

Furthermore, it can be shown that  $\widehat{\beta}_{dr}^{(1)}$  is in large samples normally distributed with mean  $\beta_0^{(1)}$  and variance that is easily estimated via the nonparametric bootstrap. The bootstrap is required here to appropriately account for additional variability from the first stage regression of  $X_i^{(1)}$  onto  $X_i^{(-1)}$ . Although doubly robust estimators of a multiplicative exposure effect have previously been proposed (Robins and Rotnitzky, 2001), the doubly robust method described here is new and has the appealing property that, unlike previous methods, it does not require an estimate of the baseline risk  $\Pr(Y = 1 | X = 0)$ .

The simulation study reported in table 1 nicely illustrates the robustness property described in Result 2, as it shows in the row labelled 'Incorrect Model', that the doubly robust estimator remains unbiased when model (5) holds, even though model (4) is incorrect because in this scenario,  $Y$  is generated under a log-binomial model with event probability  $\exp([-1.5, 0.3, -0.2, -0.7, 0.9] \times Q)$ , with corresponding marginal risk  $\Pr(Y = 1) \approx 0.30$ . This is in stark contrast with the non-doubly robust estimator  $\widehat{\beta}^{(1)}$  which incurs bias when the confounders are mis-specified. The simulation study also indicates that when modeling error is absent, the doubly robust estimator exhibits similar efficiency as the non-doubly robust estimator, suggesting that, at least in this specific simulation study, little efficiency loss was incurred in exchange for a potential gain in robustness. In

the appendix, the doubly robust methods described above are extended to incorporate possible interactions between the exposure and covariates, and the approach is further developed for a continuous exposure.

## 4 Conclusion

In this paper, we have described a simple and efficient two-stage approach to estimate risk ratios directly, which does not directly rely for consistency on an estimate of the baseline risk. This latter property is advantageous, because unlike previous methods, the proposed approach obviates the need to restrict the predicted risk probabilities to fall below one, in order to recover efficient inferences about risk ratios. For efficiency, the approach incorporates an individual weight which does depend on the individual's predicted risk; nonetheless, because the primary target of inference is the risk ratio parameter, we have argue that a consistent estimate of the risk is sufficient for inference, and we have described a simple pluggin estimator of risk which we have used to construct an efficient estimator of risk ratios. Both a simulation study and a data application confirmed the good performance of the approach. We have further extended the proposed methodology by modifying it to ensure that individuals' estimated risks are genuine probabilities. Furthermore, when the primary interest is in the effect of a specific exposure, we have developed a simple doubly robust closed-form estimator for the multiplicative effect of the exposure, while adjusting for a possibly large number of confounders. In future work, we plan to further extend the methods of this paper for correlated binary outcomes as encountered in studies with repeated outcome measurements, or in studies with clustered data.

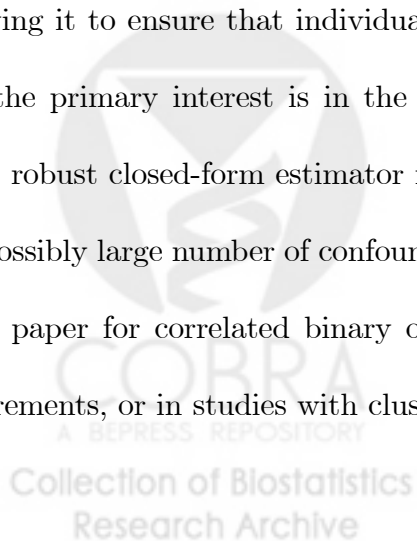


Table1. Simulation results

	bias ( $\hat{\beta}$ )	MC Var ( $\hat{\beta}$ )	$\hat{\Sigma}_{\beta}$	$\tilde{\Sigma}_{\beta}$	bias ( $\hat{\beta}^{eff}$ )	MC Var ( $\hat{\beta}^{eff}$ )	bias ( $\hat{\beta}_{dr}$ )	MC Var ( $\hat{\beta}_{dr}$ )
$\beta^{(1)}$ Correct Model	-0.0083	0.0140	0.0180	0.0138	0.0029	0.0123	-0.0079	0.0141
$\beta^{(1)}$ Incorrect Model	0.0183	0.0128	0.0170	0.0130	-0.0226	0.0110	0.0001	0.0130
$\beta^{(2)}$ Correct Model	-0.0026	0.0138	0.0185	0.0135	0.0046	0.0115	**	**
$\beta^{(3)}$ Correct Model	0.0150	0.0402	0.0514	0.0388	0.0007	0.0316	**	**
$\beta^{(4)}$ Correct Model	-0.0084	0.0411	0.0515	0.0390	-0.0022	0.0345	**	**



# APPENDIX 1

## Proofs

Proof of Result 1: Let  $E \{U^*(\beta)\} = E [Y \{\exp(-\beta^T (X - E(X)))\} (X - E(X))]$  denote the (probability) limiting value of  $n^{-1} \sum_{i:Y_i=1} U_i(\beta)$ . To show that the result holds, it suffices to show that  $U(\beta)$  is an unbiased estimating function; that is we need to show that  $E \{U(\beta_0)\} = 0$ . Now

$$\begin{aligned}
 E \{U(\beta_0)\} &= E [Y \{\exp(-\beta_0^T (X - E(X)))\} (X - E(X))] \\
 &= E [E(Y|X) \{\exp(-\beta_0^T (X - E(X)))\} (X - E(X))] \\
 &= E [\exp(\beta_0^T X + \alpha_0) \{\exp(-\beta_0^T (X - E(X)))\} (X - E(X))] \\
 &= E [\exp(\alpha_0 + \beta_0^T E(X)) (X - E(X))] \\
 &= \exp(\alpha_0 + \beta_0^T E(X)) E[(X - E(X))] \\
 &= 0
 \end{aligned}$$

To establish the large sample behaviour of  $\hat{\beta}$ , we perform a standard Taylor expansion

$$\begin{aligned}
 0 &= \sum_i [Y_i \{\exp(-\hat{\beta}^T (X_i - \bar{X}))\} (X_i - \bar{X})] \\
 &\approx \sum_i [Y_i \{\exp(-\beta_0^T (X_i - E(X)))\} (X_i - E(X))] \\
 &\quad - \sum_i [Y_i \{\exp(-\beta_0^T (X_i - E(X)))\} (X_i - E(X)) (X_i - E(X))^T] (\hat{\beta} - \beta_0) \\
 &\quad + \sum_i [Y_i \{\exp(-\beta_0^T (X_i - E(X)))\} (X_i - E(X)) \beta_0^T] (\bar{X} - E(X)) \\
 &\quad - \sum_i [Y_i \{\exp(-\beta_0^T (X_i - E(X)))\}] (\bar{X} - E(X))
 \end{aligned}$$

By the law of large numbers and an application of Slutsky's theorem, we conclude that  $\sqrt{n}(\hat{\beta} - \beta_0)$  has large sample distribution equal to the distribution of

$$\begin{aligned}
& E \left[ Y_i \{ \exp(-\beta_0^T (X - E(X))) \} (X - E(X)) (X - E(X))^T \right]^{-1} \\
& \times \left\{ n^{-1/2} \sum_i [Y_i \{ \exp(-\beta_0^T (X_i - E(X))) \} (X_i - E(X))] \right. \\
& + E [Y \{ \exp(-\beta_0^T (X - E(X))) \} (X - E(X)) \beta_0^T] n^{-1/2} \sum_i (X_i - E(X)) \\
& \left. - E [Y \{ \exp(-\beta_0^T (X - E(X))) \}] n^{-1/2} \sum_i (X_i - E(X)) \right\} \\
& = E \left[ Y_i \{ \exp(-\beta_0^T (X - E(X))) \} (X - E(X)) (X - E(X))^T \right]^{-1} \times \\
& \left\{ \sum_i [Y_i \{ \exp(-\beta_0^T (X_i - E(X))) \} - E [Y \{ \exp(-\beta_0^T (X - E(X))) \}]] (X_i - E(X)) \right\}
\end{aligned}$$

since  $E [Y \{ \exp(-\beta_0^T (X - E(X))) \} (X - E(X)) \beta_0^T] = 0$ . We may further conclude that the large sample variance of  $\sqrt{n}(\hat{\beta} - \beta_0)$  is given by

$$\begin{aligned}
& E \left[ Y_i \{ \exp(-\beta_0^T (X - E(X))) \} (X - E(X)) (X - E(X))^T \right]^{-1} \\
& \times E \left\{ [Y_i \{ \exp(-\beta_0^T (X_i - E(X))) \} - E [Y \{ \exp(-\beta_0^T (X - E(X))) \}]] (X_i - E(X)) \right\}^{\otimes 2} \\
& \times E \left[ Y_i \{ \exp(-\beta_0^T (X - E(X))) \} (X - E(X)) (X - E(X))^T \right]^{-1} \\
& = \left[ E \left( \frac{\partial U^*(\beta)}{\partial \beta} \right) \right]^{-1} E \{ U^*(\beta_0)^{\otimes 2} \} \left[ E \left( \frac{\partial U^*(\beta)}{\partial \beta} \right) \right]^{-1} \\
& - E [Y \{ \exp(-\beta_0^T (X - E(X))) \}]^2 \\
& \times \left[ E \left( \frac{\partial U^*(\beta)}{\partial \beta} \right) \right]^{-1} E \{ (X - E(X))^{\otimes 2} \} \left[ E \left( \frac{\partial U^*(\beta)}{\partial \beta} \right) \right]^{-1}
\end{aligned}$$

because

$$\begin{aligned} & E \left[ Y \left\{ \exp(-\beta_0^T (X - E(X))) \right\} \right] \\ &= E \left[ Y \left\{ \exp(-\beta_0^T (X - E(X))) \right\} (X - E(X)) (X - E(X))^T \right] E \left[ (X - E(X)) (X - E(X))^T \right]^{-1} \end{aligned}$$

where  $A^{\otimes 2} = AA^T$ . Furthermore, because covariance matrices are positive-definite, we may conclude that  $\left[ E \left( \frac{\partial U^*(\beta)}{\partial \beta} \right) \right]^{-1} E \{ U^*(\beta_0)^{\otimes 2} \} \left[ E \left( \frac{\partial U^*(\beta)}{\partial \beta} \right) \right]^{-1}$  is conservative for the variance-covariance matrix in the positive-definite sense, that for any non-zero constant vector  $t$

$$t^T \Sigma_\beta t < t^T \left[ E \left( \frac{\partial U^*(\beta)}{\partial \beta} \right) \right]^{-1} E \{ U^*(\beta_0)^{\otimes 2} \} \left[ E \left( \frac{\partial U^*(\beta)}{\partial \beta} \right) \right]^{-1} t$$

and therefore  $\widehat{\Sigma}_\beta$  is a conservative estimator of  $\Sigma_\beta$ . Whereas  $\widetilde{\Sigma}_\beta$  is consistent for  $\Sigma_\beta$  where

$$\begin{aligned} \widetilde{\Sigma}_\beta &= \widehat{\Sigma}_\beta - \left[ \sum_i Y_i \left\{ \exp(-\beta_0^T (X_i - \bar{X})) \right\} \right]^2 \\ &\times \left[ \sum_{i:Y_i=1} \frac{\partial U_i(\beta)}{\partial \beta} \Big|_{\widehat{\beta}} \right]^{-1} n \sum_{i:Y_i=1} (X_i - \bar{X})^{\otimes 2} \left[ \sum_{i:Y_i=1} \frac{\partial U_i(\beta)}{\partial \beta} \Big|_{\widehat{\beta}} \right]^{-1} \end{aligned}$$

Proof of Result 2: Consider the semiparametric model given solely by restriction (3); then Bickel et al (1993) established that all regular and asymptotically linear estimators of  $\beta_0$  are fully characterized by the set of influence functions :

$$\Lambda = \left\{ \begin{array}{l} U^\dagger(v) = \left\{ Y - \exp(\alpha_0 + \beta_0^T X) \right\} \times \left[ V - \frac{E\{V \exp(\beta_0^T X)\}}{E\{\exp(\beta_0^T X)\}} \right] : \\ V = v(X) \text{ of dimension } \dim -\beta_0 \\ \text{with } E \left\{ U^\dagger(v)^T U^\dagger(v) \right\} < \infty \end{array} \right\}$$

It is straightforward to verify that this set is equivalently written :

$$\Lambda = \left\{ \begin{array}{l} U^\natural(\mu) = \frac{\{Y - \exp(\alpha_0 + \beta_0^T X)\}}{1 - \exp(\alpha_0 + \beta_0^T X)} \times \left[ \mu - \frac{E\{\mu \exp(\alpha_0 + \beta_0^T X) [1 - \exp(\alpha_0 + \beta_0^T X)]^{-1}\}}{E\{\exp(\alpha_0 + \beta_0^T X) [1 - \exp(\alpha_0 + \beta_0^T X)]^{-1}\}} \right] : \\ \mu = \mu(X) \text{ of dimension } \dim -\beta_0 \\ \text{with } E\{U^\natural(\mu)^T U^\natural(\mu)\} < \infty \end{array} \right\}$$

Now, the score for  $\beta_0$  in this model is given by

$$S_\beta = \frac{X(Y - \exp(\alpha_0 + \beta_0^T X))}{\{1 - \exp(\alpha_0 + \beta_0^T X)\}}$$

therefore, the efficient score of  $\beta_0$ , i.e. the orthogonal projection of  $S_\beta$  onto  $\Lambda$ , is  $U^\dagger(\mu_{opt})$ , with

$\mu_{opt} = \mu_{opt}(X) = X$ , in other words,

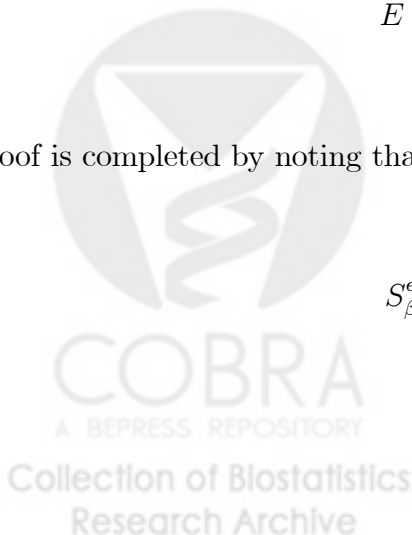
$$S_\beta^{eff} = U^\natural(\mu_{opt}) = \frac{\{Y - \exp(\alpha_0 + \beta_0^T X)\}}{1 - \exp(\alpha_0 + \beta_0^T X)} \times \left[ X - \frac{E\{X \exp(\alpha_0 + \beta_0^T X) [1 - \exp(\alpha_0 + \beta_0^T X)]^{-1}\}}{E\{\exp(\alpha_0 + \beta_0^T X) [1 - \exp(\alpha_0 + \beta_0^T X)]^{-1}\}} \right]$$

since  $S_\beta^{eff} \in \Lambda$ , and for all  $U^\natural(w) \in \Lambda$

$$E\left[\left(S_\beta - S_\beta^{eff}\right) U^\natural(w)\right] = 0$$

The proof is completed by noting that

$$S_\beta^{eff} = U^\natural(\mu_{opt}) = U^\dagger(v_{opt})$$



where

$$v_{opt}(X) = \left[ X - \frac{E \left\{ X \exp(\alpha_0 + \beta_0^T X) [1 - \exp(\alpha_0 + \beta_0^T X)]^{-1} \right\}}{E \left\{ \exp(\alpha_0 + \beta_0^T X) [1 - \exp(\alpha_0 + \beta_0^T X)]^{-1} \right\}} \right] / 1 - \exp(\alpha_0 + \beta_0^T X)$$

Then, a theorem due to Bickel et al (1993) states that for any initial  $n^{1/2}$ -consistent estimator of  $\beta_0$ , an efficient estimator can be constructed by a one-step update of  $\hat{\beta}$  in the direction of the estimated efficient score by using the following formula

$$\hat{\beta}^{eff} = \hat{\beta} - \left[ \sum_i \hat{U}_i^{\cdot \natural}(\mu_{opt}) \right]^{-1} \sum_i \hat{U}_i^{\natural}(\mu_{opt})$$

where  $\hat{U}^{\natural}(\mu_{opt})$  is an empirical version of  $U^{\natural}(\mu_{opt})$  obtained by replacing all expectations by empirical expectations, with  $\beta_0$  estimated by  $\hat{\beta}$  and  $\exp(\alpha_0)$  estimated by the simple pluggin estimator  $\sum_{i'} Y_{i'} \exp(-\hat{\beta}^T X_{i'})/n$ ;  $\sum_i \hat{U}^{\cdot \natural}(\mu_{opt})/n$  is a similarly constructed estimator of the expected derivative of the efficient score,  $\frac{\{Y - \exp(\alpha_0 + \beta^T X)\}}{1 - \exp(\alpha_0 + \beta^T X)} \left[ X - \frac{E\{X \exp(\alpha_0 + \beta^T X) [1 - \exp(\alpha_0 + \beta^T X)]^{-1}\}}{E\{\exp(\alpha_0 + \beta^T X) [1 - \exp(\alpha_0 + \beta^T X)]^{-1}\}} \right]$  with respect to  $\beta$  evaluated at  $\beta_0$ . It is straightforward to verify that  $\hat{\beta}^{eff}$  reduces to the formula provided in the main text. Furthermore, the theorem of Bickel et al (1993) further states that under standard regularity conditions,  $n^{1/2}(\hat{\beta}^{eff} - \beta_0)$  is asymptotically normal with mean zero and variance

$$E \{ U^{\natural}(\mu_{opt})^{\otimes 2} \}^{-1}$$

which is also the semiparametric efficiency bound of  $\beta_0$ . Finally,  $\hat{\Sigma}_{\beta}^{eff}$  is an empirical version of  $\Sigma_{\beta}^{eff}$  which converges to the latter in probability.

In order to prove Result 3, we first establish a more general result, for which we allow  $X^{(1)}$  to be continuous, and for the model to incorporate a possible interaction between exposure and

covariates, say  $X^{(2)}$ . Specifically, we suppose that

$$\frac{\Pr(Y_i = 1|X^{(1)}, X^{(-1)})}{\Pr(Y_i = 1|X^{(1)} = 0, X^{(-1)})} = \exp \left\{ \beta_0^{(1)T} \left[ X_i, X_i^{(1)} X_i^{(2)} \right]^T \right\} \quad (6)$$

and let  $\pi(\psi) = E(X^{(1)}|X^{(-1)}; \psi) = g(\psi^T[1, X^{(-1)T}]^T)$  denote a working model for the mean of the exposure given covariates; where  $g^{-1}$  is the identity link for continuous  $X^{(1)}$  and  $g$  is the logit link for binary  $X^{(-1)}$ . Define the estimating function

$$W(\beta^{(1)}, \beta^{(-1)}, \psi) = \begin{pmatrix} 1 \\ X^{(2)} \end{pmatrix} Y_i \exp \left\{ -\beta^{(1)T} [X, X^{(1)} X^{(2)}]^T - \beta^{(-1)T} X^{(-1)} \right\} \{X^{(1)} - \pi(\psi)\}$$

Then we have the following lemma.

*Lemma 1: Under model (6),*

$$E \left[ W(\beta_0^{(1)}, \beta^{(-1)}, \psi) \right] = 0 \quad (7)$$

*if either but not necessarily both of the following conditions hold,*

- (1)  $\psi = \psi_0$  and  $E(X^{(1)}|X^{(-1)}; \psi_0) = E(X^{(1)}|X^{(-1)})$  or
- (2)  $\beta^{(-1)} = \beta_0^{(-1)}$  and model (4) holds.



Proof of Lemma 1:

$$\begin{aligned}
 E \left[ W(\beta_0^{(1)}, \beta^{(-1)}, \psi) \right] &= E \left[ \begin{pmatrix} 1 \\ X^{(2)} \end{pmatrix} Y_i \exp \left\{ -\beta_0^{(1)T} [X, X^{(1)} X^{(2)}]^T - \beta^{(-1)T} X^{(-1)} \right\} \{ X^{(1)} - \pi(\psi) \} \right] \\
 &= E \left[ \begin{pmatrix} 1 \\ X^{(2)} \end{pmatrix} E[Y_i | X_i] \exp \left\{ -\beta_0^{(1)T} [X, X^{(1)} X^{(2)}]^T - \beta^{(-1)T} X^{(-1)} \right\} \{ X^{(1)} - \pi(\psi) \} \right] \\
 &= E \left[ \begin{pmatrix} 1 \\ X^{(2)} \end{pmatrix} \exp \left\{ E[Y_i | X_i^{(1)} = 0, X_i^{(-1)}] - \beta^{(-1)T} X^{(-1)} \right\} \{ E(X^{(1)} | X^{(-1)}) - \pi(\psi) \} \right]
 \end{aligned}$$

which is certainly zero if (1) holds since then  $E(X^{(1)} | X^{(-1)}) - \pi(\psi) = 0$ . If (2) holds, we have

$$\begin{aligned}
 &E \left[ \begin{pmatrix} 1 \\ X^{(2)} \end{pmatrix} \exp \left\{ E[Y_i | X_i^{(1)} = 0, X_i^{(-1)}] - \beta^{(-1)T} X^{(-1)} \right\} \{ E(X^{(1)} | X^{(-1)}) - \pi(\psi) \} \right] \\
 &= E \left[ \begin{pmatrix} 1 \\ X^{(2)} \end{pmatrix} \exp \{ E[Y_i | X_i = 0] \} \{ E(X^{(1)} | X^{(-1)}) - \pi(\psi) \} \right] \\
 &= \exp \{ E[Y_i | X_i = 0] \} \times E \left[ \begin{pmatrix} 1 \\ X^{(2)} \end{pmatrix} \{ X^{(1)} - \pi(\psi) \} \right] \\
 &\propto E \left[ \begin{pmatrix} 1 \\ X^{(2)} \end{pmatrix} \{ X^{(1)} - \pi(\psi) \} \right] = 0
 \end{aligned}$$

since the last quantity is part of the first order condition used to estimate  $\psi$  either by ordinary least-squares when  $X^{(-1)}$  is continuous or by logistic regression in the binary case.

Proof of Result 3: The result immediately follows from Lemma 1 since when  $X^{(1)}$  is binary, it is straightforward to verify that equation (7) is equivalent to

$$\beta_0^{(1)} = \log \frac{E [Y X^{(1)} \exp(-\beta^{(-1)T} X^{(-1)}) \{1 - \pi(\psi)\}]}{E [Y (1 - X^{(1)}) \exp(-\beta^{(-1)T} X^{(-1)}) \pi(\psi)]}$$

Therefore, if either (1) holds, and thus  $\hat{\psi}$  converges to  $\psi_0$  or (2) holds and thus  $\hat{\beta}^{(-1)}$  converges to  $\beta_0^{(-1)}$ , we have that  $\hat{\beta}_{dr}^{(1)}$  converges to  $\beta_0^{(1)}$ .

## References

- [1] Bickel, P., Klassen, C., Ritov, Y. and Wellner, J. (1993). Efficient and Adaptive Estimation for Semi-parametric Models. Springer, New York.
- [2] Breslow N E. Covariance analysis of censored survival data. Biometrics 1974; 30: 89-99.
- [3] Chu H, Cole S. Estimation of Risk Ratios in Cohort Studies With Common Outcomes. A Bayesian Approach. Epidemiology (2010) Volume 21, Number 6.
- [4] Deddens JA, Petersen MR, Lei X. (2003) Estimation of prevalence ratios when PROC GENMOD does not converge. Proceedings of the 28th Annual SAS Users Group International Conference, Seattle, Washington, Paper 270-28. (<http://www2.sas.com/proceedings/sugi28/270-28.pdf>).
- [5] The Diabetes Control and Complications Trial Research Group (1993) The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med :329:977–86.



- [6] Greenland S.(2004) Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. *Am J Epidemiol*;160:301–305.
- [7] Hirano, K., Imbens, G.W., Ridder, G. (2003) Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71, 1161–1189.
- [8] Lachin J. (2000). *Biostatistical Methods*. John Wiley and Sons. New York.
- [9] Hastie T, Tibshirani R and Friedman J (2008). *The Elements of Statistical Learning* (2nd edition).Springer-Verlag, New York.
- [10] Lee, J. Odds Ratio or Relative Risk for Cross-Sectional Data? *Int. J. Epidemiol.* (1994) 23(1): 201-203 doi:10.1093/ije/23.1.20
- [11] Petersen MR, Deddens JA.(2006) RE: “Easy SAS calculations for risk or prevalence ratios and differences.” *Am J Epidemiol*;163:1158–1159.
- [12] Ruppert D, Wand M, Carroll R (2003) *Semiparametric Regression*. Cambridge University Press.
- [13] Robins JM, Rotnitzky A. (2001). Comment on the Bickel and Kwon article, "Inference for semiparametric models: Some questions and an answer" *Statistica Sinica*, 11(4):920-936.
- [14] Skov T, Deddens J, Petersen MR, Endahl L. (1998) Prevalence proportion ratios: estimation and hypothesis testing. *Int J Epidemiol*;27:91–95.
- [15] Spiegelman D, Hertzmark E.(2005) Easy SAS calculations for risk or prevalence ratios and differences. *Am J Epidemiol*;162:199 –200.
- [16] Tian L, Liu, K.(2006) Re: “Estimating the relative risk in cohort studies and clinical trials of common outcomes.” (Letter). *Am J Epidemiol*;163:1157–1163.

- [17] Wacholder S.(1986) Binomial regression in GLIM: estimating risk ratios and risk differences. Am J Epidemiol;123:174 –184.
- [18] Wasserman L. (2003). All of Nonparametric Statistics. Springer, New York.
- [19] Zou GY.(2004) A modified Poisson regression approach to prospective studies with binary data. Am J Epidemiol.;159:702–706.

