Published in final edited form as:

J Stat Plan Inference. 2012 December 1; 142(12): 3235–3241. doi:10.1016/j.jspi.2012.05.005.

Reducing Bias and Mean Squared Error Associated With Regression-Based Odds Ratio Estimators

Robert H. Lyles^a, Ying Guo^a, and Sander Greenland^b

^aDepartment of Biostatistics and Bioinformatics, The Rollins School of Public Health of Emory University, 1518 Clifton Rd. N.E., Atlanta, GA 30322 phone: 404-727-1310; fax: 404-727-1370

^bDepartment of Epidemiology and Department of Statistics, University of California at Los Angeles, Los Angeles, CA 90095-1772

Abstract

Ratio estimators of effect are ordinarily obtained by exponentiating maximum-likelihood estimators (MLEs) of log-linear or logistic regression coefficients. These estimators can display marked positive finite-sample bias, however. We propose a simple correction that removes a substantial portion of the bias due to exponentiation. By combining this correction with bias correction on the log scale, we demonstrate that one achieves complete removal of second-order bias in odds ratio estimators in important special cases. We show how this approach extends to address bias in odds or risk ratio estimators in many common regression settings. We also propose a class of estimators that provide reduced mean bias and squared error, while allowing the investigator to control the risk of underestimating the true ratio parameter. We present simulation studies in which the proposed estimators are shown to exhibit considerable reduction in bias, variance, and mean squared error compared to MLEs. Bootstrapping provides further improvement, including narrower confidence intervals without sacrificing coverage.

Keywords

Absolute risk; Bias; Bootstrap; Logistic regression; Maximum likelihood; Odds ratio; Relative risk

1. INTRODUCTION

Generalized linear models yield fitted coefficients that are commonly used to estimate odds ratios or other measures of association. Standard fitting techniques such as maximum-likelihood and estimating equation methods yield consistent estimators with first-order asymptotically normal sampling distributions (Cox and Oakes 1984; McCullagh and Nelder 1989; Davidian and Giltinan 1995; Diggle et al. 2002; Agresti 2002). Outside of linear models, however, these estimators can suffer from considerable higher-order asymptotic bias.

Most research on bias reduction has targeted the fitted regression coefficients, e.g., Byth and McLachlan (1978), Anderson and Richardson (1979), McLachlan (1980), Schaeffer (1983),

© 2012 Elsevier B.V. All rights reserved.

Corresponding author: rlyles@sph.emory.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Cook, Tsai, and Wei (1986), Cordeiro and McCullagh (1991), and Firth (1993), although King and Zeng (2001) studied bias-reduced estimators for probabilities based on logistic regression with rare events. While some statisticians see arithmetic bias and mean squared error (MSE) as less relevant on the scale of skewed estimators of effect (e.g., odds ratios), this is not a universal view. Several authors (e.g., Jewell 1984, 1986; Greenland 2000) have targeted small-sample bias on the odds ratio scale in tabular-based analyses, noting that researchers generally report and interpret point estimates on that scale. Such reporting has added practical rationale in that for uncommon diseases, the odds ratio approximates the risk ratio and thus is proportional to the change in caseload associated with a unit increase in exposure (Greenland, 2000).

We propose an approach to bias reduction for estimating measures of effect in general regression settings. Exponentiation of estimated regression coefficients to transform to an absolute-effect scale induces positive bias, a well-known consequence of Jensen's inequality (Jensen 1906). We derive a form for this positive second-order bias that leads to a reduced-bias estimator. We then explore further benefits of an initial first-order bias correction on the log scale. Our main estimator targets second-order unbiasedness at the expense of increasing median bias (Read, 1985) relative to standard estimators. We propose alternative bias-reduced estimators to control the increased risk of underestimation due to mean bias adjustment.

Throughout, we use “asymptotic” and “approximate” to indicate standard first-order results (which apply on a scale proportional to $n^{-1/2}$ where n is the sample size) and will contrast those to second-order (scale n^{-1}) properties, comparing finite-sample behavior via simulations.

2. METHODS AND EXAMPLES

2.1 Bias-Corrected Odds Ratio Estimation

Odds ratio (OR) and rate or risk ratio (RR) estimators typically arise from exponentiating coefficient estimators from logistic, Poisson, or Cox models (Cox and Oakes 1984; McCullagh and Nelder 1989; Agresti 2002). Consider a generalized linear model of the form

$$g[E(Y|\mathbf{X}=\mathbf{x})]=\beta_0+\sum_{j=1}^k\beta_jx_j, \quad (1)$$

where $g(\cdot)$ is a strictly increasing link function. Inferences based on (1) commonly make use of the asymptotic normality of maximum-likelihood (ML) or other standard estimators:

$$\hat{\beta}_j \overset{\bullet}{\sim} N(\beta_j, \sigma_j^2), \quad (2)$$

where σ_j^2 is the asymptotic variance of $\hat{\beta}_j$. Thus, $\hat{\beta}_j$ is asymptotically median unbiased.

Typically, the link function g is the logit or logarithm, and the target parameter is $\psi_j = e^{\beta_j}$, with estimator $\hat{\psi}_j = e^{\hat{\beta}_j}$ ($j=1, \dots, k$). Although the first-order limiting distribution of $\hat{\psi}_j$ is also normal with mean ψ_j , the log-scale normal approximation is far more accurate for typical sample sizes. It follows from (2) that the distribution of $\hat{\psi}_j$ will be more closely lognormal with

$$E(\widehat{\Psi}_j) \doteq e^{\beta_j + \sigma_j^2/2} = e^{\sigma_j^2/2} \psi_j. \quad (3)$$

Expression (3) may also be derived via Taylor-series arguments given that $E(\hat{\beta}_j) \approx \beta_j$, without assuming normality for $\hat{\beta}_j$. The bias factor $e^{\sigma_j^2/2}$ disappears asymptotically.

Approximate median unbiasedness of $\hat{\beta}_j$ for β_j ensures approximate median unbiasedness of $\widehat{\Psi}_j$ for ψ_j . However, $\widehat{\Psi}_j$ can be subject to large overestimation errors unless σ_j^2 is small. Such errors might be especially detrimental if the estimate is used for resource-allocation decisions, e.g., if it is taken as representing the best prediction of the excess caseload to be expected from a harmful ($\psi_j > 1$) exposure (e.g., Greenland, 2000). The bias factor ($e^{\sigma_j^2/2}$) is negligible for small but increases rapidly with σ_j : The approximate expectation of $\widehat{\Psi}_j$ overestimates ψ_j by more than 50% for σ_j values > 0.9 (see Supplementary Figure 1).

To reduce the bias of $\widehat{\Psi}_j$, consider the following corrected estimator based on eqn. (3):

$$\widehat{\Psi}_{j,\text{corr}} = e^{-\widehat{\sigma}_j^2/2} \widehat{\Psi}_j, \quad (4)$$

where $\widehat{\sigma}_j^2$ is the variance estimate for $\hat{\beta}_j$. This “plug-in” estimator is necessarily smaller than $\widehat{\Psi}_j$. A criticism of (4) is that it only reduces bias due to exponentiation, without addressing higher-order bias in the log-scale estimator $\hat{\beta}_j$. We will investigate the extent to which initial bias correction to $\hat{\beta}_j$ followed by (4) improves performance.

2.2 Bias-Reduced OR Estimators Controlling the Risk of Underestimation

The estimator in (4) designed to reduce mean bias is first-order equivalent to the MLE, but has lower second-order variance due to multiplying by a correction factor between 0 and 1. As it converges more slowly to median unbiasedness, we also consider estimators that compromise between $\widehat{\Psi}_j$ and $\widehat{\Psi}_{j,\text{corr}}$. To this end, note that (2) implies that

$$\Pr(\widehat{\Psi}_{j,\text{corr}} < e^{\beta_j}) \doteq \Phi(\sigma_j/2), \quad (5)$$

where $\Phi(\cdot)$ is the standard normal CDF. Thus $\Phi(\widehat{\sigma}_j/2)$, which always exceeds 0.5, estimates the probability that $\widehat{\Psi}_{j,\text{corr}}$ would underestimate ψ_j . To reduce median bias and target a risk of underestimation no larger than p (< 0.5), we consider estimators of the form $\widehat{\Psi}_{j,p} = e^{c\widehat{\Psi}_j}$ for some constant c such that $\Pr(\widehat{\Psi}_{j,p} < \Psi_j) \approx p$. From eqn. (2), the latter relation holds when $c = -\sigma_j z_p$, where z_p is the 100 $\times p$ -th percentile of the standard normal distribution. This leads to a class of estimators that are consistent for all p , i.e.,

$$\widehat{\Psi}_{j,p} = e^{-\widehat{\sigma}_j z_p} \widehat{\Psi}_j. \quad (6)$$

To improve mean bias and squared error while limiting the increased chance of underestimating Ψ_j in repeated samples, we propose the following bias-reduced estimator:

$$\widehat{\Psi}_{j,\text{corr}}^* = \max(\widehat{\Psi}_{j,\text{corr}}, \widehat{\Psi}_{j,p}), \quad (7)$$

with p selected by the investigator. This encompasses two extremes: 1) An investigator who demands median unbiasedness takes $p = 0.5$, so $\widehat{\Psi}_{j,\text{corr}}^* = \widehat{\Psi}_{j,.50} = \widehat{\Psi}_j$, the usual MLE; 2) An investigator unconcerned with median unbiasedness uses $\widehat{\Psi}_{j,\text{corr}}$. Note that (7) is equivalent to

$$\hat{\Psi}_{j,\text{corr}}^* = \begin{cases} \hat{\Psi}_{j,\text{corr}} & \text{if } \hat{\sigma}_j \leq 2z_p \\ \hat{\Psi}_{j,p} & \text{otherwise} \end{cases}$$

Thus, the bias-reduced estimator retains the full mean-bias correction as in (4) when $\hat{\sigma}_j \leq 2z_p$; otherwise, it reduces the correction factor to a degree determined by p , the acceptable probability of underestimation. In Section 3, we summarize a simulation study to evaluate the performance of a version of (7) in which a bias correction is initially applied to $\hat{\beta}_j$.

2.3 Case Studies 1 and 2: 2×2 Tables for Unpaired and Paired Data

For brevity, we confine attention to odds ratios. To investigate the bias correction (4), let X and Y respectively denote exposure and disease status and consider 2×2 tables for unadjusted odds-ratio estimation in unpaired and pair-matched case-control settings as in Table 1.

The number of subjects in the unpaired setting is $n = A+B+C+D$, whereas for paired data n is the number of pairs. Under Poisson, product-binomial and multinomial models for the unpaired data cell counts, the MLE for the odds ratio $\psi = p_{11}p_{00}/(p_{10}p_{01})$ in Table 1 is $\hat{\Psi}_{\text{ML}} = AD/BC = \hat{p}_{11}\hat{p}_{00}/(\hat{p}_{10}\hat{p}_{01})$, and $\hat{\sigma}^2 = 1/A+1/B+1/C+1/D$ is the usual first-order variance estimator for $\hat{\beta}_{\text{ML}} = \ln(\hat{\Psi}_{\text{ML}})$. Under Poisson and multinomial models for the pair counts, the MLE for the paired-data odds ratio $\psi = p_{10}/p_{01}$ in Table 1 is $\hat{\Psi}_{\text{ML}} = B/C = \hat{p}_{10}/\hat{p}_{01}$, and $\hat{\sigma}^2 = 1/B+1/C$ is the usual first-order variance estimator for $\hat{\beta}_{\text{ML}}$.

The $O(n^{-1})$ bias in $\hat{\Psi}_{\text{ML}}$ and $\hat{\beta}_{\text{ML}} = \ln(\hat{\Psi}_{\text{ML}})$ follows from the Taylor-series expansion

$$E[f(\hat{\mathbf{p}})] = f(\mathbf{p}) + g(\mathbf{p}) + O(n^{-2}), \quad (8)$$

where $g(\mathbf{p}) = E[(\hat{\mathbf{p}} - \mathbf{p})' D_2(\mathbf{p})(\hat{\mathbf{p}} - \mathbf{p})/2]$ and $D_2(\mathbf{p})$ is the Hessian of f evaluated at $\mathbf{p}$ (Jewell 1984). Along with the previous estimator $\hat{\Psi}_{\text{corr}1} = e^{-\hat{\sigma}^2/2} \hat{\Psi}_{\text{ML}}$ (we now add the subscript 1), we consider $\hat{\Psi}_{\text{corr}2} = e^{-\hat{\sigma}^2/2} \exp(\hat{\beta}_{\text{ML}}^*)$ where $\hat{\beta}_{\text{ML}}^*$ is the bias-corrected estimator for $\beta = \ln(\text{OR})$ derived from (8). For unpaired data, $\hat{\beta}_{\text{ML}}^* = \hat{\beta}_{\text{ML}} - (1/2)(1/B+1/C - 1/A-1/D)$, while for paired data $\hat{\beta}_{\text{ML}}^* = \hat{\beta}_{\text{ML}} - (1/2)(1/C-1/B)$. Table 2 summarizes the resulting $O(n^{-1})$ bias terms.

In each case the arithmetic bias term for $\hat{\Psi}_{\text{ML}}$ is positive, and could be quite large. The $O(n^{-1})$ bias of $\hat{\Psi}_{\text{corr}1}$ is not null, but tends to be markedly smaller than that of $\hat{\Psi}_{\text{ML}}$ when $\psi > 1$. Most importantly, implementing an $O(n^{-1})$ bias correction to the log-scale estimator $\hat{\beta}_{\text{ML}}$ before applying the correction factor $e^{-\hat{\sigma}^2/2}$ to obtain $\hat{\Psi}_{\text{corr}2}$, completely eliminates the $O(n^{-1})$ bias.

The estimator $\hat{\Psi}_{\text{corr}2}$ eliminates $O(n^{-1})$ bias even though it uses the first-order variance estimator ($\hat{\sigma}^2$) for $\hat{\beta}_{\text{ML}}$ in the correction factor $e^{-\hat{\sigma}^2/2}$. One might expect small-sample performance improvements by using an alternative estimator for σ^2 that reflects the (slightly reduced) variance of the bias-corrected estimator $\hat{\beta}_{\text{ML}}^*$. We explore this idea below.

2.4 Bias-Corrected OR Estimation in Logistic Regression

Our general strategy to obtain bias-corrected adjusted OR estimates begins with model (1), assuming a logit link. While direct application of eqn. (4) reduces bias, further benefits follow from initial application of an $O(n^{-1})$ coefficient-scale bias correction as available for all generalized linear models (e.g., McCullagh and Nelder 1989; Cordeiro and McCullagh 1991; Firth 1993). Because we encountered no separation problems in our simulations (see

Discussion), we focus here on the direct method of Cordeiro and McCullagh (1991) which yields the following $O(n^{-1})$ bias corrected estimator for the vector of logistic regression coefficients β :

$$\hat{\beta}^* = \hat{\beta}_{ML} - \hat{b}_1/n = \hat{\beta}_{ML} - (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\xi. \quad (9)$$

In (9), i indexes the observation and π_i the estimated binomial probability for observation i , $\mathbf{W}$ is the diagonal matrix of binomial variances $n_i\pi_i(1-\pi_i)$, and $\mathbf{W}\xi$ is a vector with $\xi_i = h_i(\pi_i^{-1/2})$ and h_i the i -th diagonal element of $\mathbf{W}^{1/2}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{1/2}$. We apply the bias-corrected estimator in (9) and then the exponentiation correction (4). For the variance estimate ($\hat{\sigma}_j^2$) to accompany each $\hat{\beta}_j^*$, there are several options. Using the variance associated with the MLE is justifiable asymptotically (Firth, 1993). However, the bias-corrected $\hat{\beta}_j^*$'s tend to shrink toward their means, resulting in finite-sample variance smaller than that of the MLE. Thus, in Section 3 we also evaluate the use of bootstrap variance estimates in (4).

2.5 Interval Estimation and Invariance

Our primary focus is point estimation on the OR scale, whereas Wald-type confidence intervals (CIs) are best set on the log scale. Thus, for CIs to accompany the bias-corrected OR estimates, we favor use of the exponentiated limits of Wald-type intervals based on bootstrap estimates of standard errors for bias-corrected coefficient estimates. Potential benefits of such intervals relative to standard ML-based alternatives are illustrated via simulations below.

If a covariate X_j is binary, then the MLE ($\hat{\Psi}_j$) for the OR has inversion symmetry. For example, if the coding of X_j is switched from (0,1) to (1,0), then $\hat{\Psi}_j$ is correspondingly inverted. In contrast, bias corrections on the OR scale are asymmetric. As a result, the proposed bias-corrected OR estimators should *not* be inverted to estimate Ψ_j upon recoding of a covariate or outcome. Doing so yields an estimator no longer bias-corrected on the inverted scale, thus negating the purpose of the proposed estimation method. The proper approach is to compute the corrected estimate directly, after first selecting the coding and scale for reporting.

3. SIMULATION STUDIES

3.1 Bias-Corrected OR Estimator Based on Discriminant Function Approach

We first replicate a simulation study originally conducted by Lyles, Guo and Hill (2009) to evaluate a discriminant function-based estimator for β , the $\ln(\text{OR})$ relating a continuous exposure X to a binary outcome Y and covariates $\mathbf{C}$. Data were simulated under the following linear model with independent normal(0,0.04) errors:

$$E(X) = 4.52 + 0.083y + 0.11c_1 - 0.07c_2 - 0.04c_3 + 0.26c_4 + 0.01c_5.$$

The parameters and covariate distributions were chosen to mimic a birthweight data example in Hosmer and Lemeshow (2000). This special case allows a unique application of the proposed bias correction approach because a strictly unbiased estimator ($\hat{\beta}_{\text{disc}}$) for the log OR (β), as well as an unbiased estimator $\hat{V}\text{ar}(\hat{\beta}_{\text{disc}})$ for its exact sampling variance, are available (Lyles et al. 2009). The estimator $\hat{\beta}_{\text{disc}}$ showed lower variance than the MLE $\hat{\beta}_{\text{ML}}$ from logistic regression of Y on $(X, \mathbf{C})$. However, the OR estimator $\hat{\Psi}_{\text{disc}} = \exp(\hat{\beta}_{\text{disc}})$ retained substantial bias due to exponentiation despite improved bias and variance relative to the MLE.

Supplementary Table 1 summarizes a repeat of the prior simulation study adding bias-corrected OR estimators

$$\hat{\Psi}_{\text{corr } 1} = e^{-\widehat{\text{Var}}(\hat{\beta}_{\text{ML}})/2} \hat{\Psi}_{\text{ML}} \text{ and } \hat{\Psi}_{\text{corr } 2} = e^{-\widehat{\text{Var}}(\hat{\beta}_{\text{disc}})/2} \hat{\Psi}_{\text{disc}}. \quad (10)$$

Note that $\hat{\Psi}_{\text{corr } 1}$ is the estimator in (4), while $\hat{\Psi}_{\text{corr } 2}$ refines it by substituting $\hat{\beta}_{\text{disc}}$ and its unbiased variance estimator. Based on 10,000 simulations with a true adjusted OR of $\exp(0.083/0.04) = 7.96$, the estimators $\hat{\Psi}_{\text{ML}}$ (mean = 13.51) and $\hat{\Psi}_{\text{disc}}$ (mean = 11.53) display extreme positive bias. In contrast, $\hat{\Psi}_{\text{corr } 1}$ is largely corrected (mean = 8.96) and $\hat{\Psi}_{\text{corr } 2}$ in (10) is virtually unbiased (mean = 7.91), with both showing striking variance and MSE reduction.

3.2 Bias-Corrected OR Estimators in Logistic Regression

Tables 3 and 4 summarize simulations based on 10,000 independent trials from a logistic model with three covariates distributed as follows: $X_1 \sim N(0, 0.25^2)$, $X_2 \sim \text{Bernoulli}(0.30)$, and $X_3 \sim \text{Uniform}(0, 0.5)$, and sample size $n=200$. The parameters β_1 , β_2 , and β_3 were set to 1.75, 1, and -0.5 , corresponding to ORs of $\Psi_1=5.75$, $\Psi_2=2.72$, and $\Psi_3=0.61$.

Table 3 compares the coefficient MLEs ($\hat{\beta}_{\text{ML}}$) versus the bias-corrected alternatives ($\hat{\beta}_{\text{corr}}$) in (9). The latter display smaller bias and variance in each case. The ML-based standard errors match the empirical standard deviation of ($\hat{\beta}_{\text{corr}}$) relatively well on average, but noticeably better matches are obtained from a parametric bootstrap (Efron and Tibshirani 1993). Wald-type confidence intervals are evaluated based on ML and on $\hat{\beta}_{\text{corr}}$ with its bootstrap standard error; the latter CIs exhibit reduced mean width, yet retain near-nominal coverage..

Table 4 summarizes the comparison of four adjusted OR estimators: $\hat{\Psi}_{\text{ML}}$, $\hat{\Psi}_{\text{corr } 1}$, $\hat{\Psi}_{\text{corr } 2}$, and $\hat{\Psi}_{\text{corr } 2}^*$. The corrected estimators are defined as follows:

$$\hat{\Psi}_{\text{corr } 1} = e^{-\widehat{\text{Var}}(\hat{\beta}_{\text{ML}})/2} \hat{\Psi}_{\text{ML}}, \quad \hat{\Psi}_{\text{corr } 2} = e^{\widehat{\beta}_{\text{corr}} - \widehat{\text{Var}}(\hat{\beta}_{\text{corr}})/2},$$

and

$$\hat{\Psi}_{\text{corr } 2}^* = \max(\hat{\Psi}_{\text{corr } 2}, \hat{\Psi}_{\text{p,corr}}), \quad (11)$$

where $\hat{\Psi}_{\text{p,corr}} = e^{\widehat{\beta}_{\text{corr}} - z_p \sqrt{\widehat{\text{Var}}(\hat{\beta}_{\text{corr}})}}$, $\widehat{\text{Var}}(\hat{\beta}_{\text{corr}})$ is based on the parametric bootstrap, and $\hat{\Psi}_{\text{p,corr}}$ is computed taking $p=0.55$ to target a risk of underestimating the true OR no larger than 55%. Table 4 shows marked positive mean bias of the standard OR estimators, especially for those corresponding to larger sampling variances (i.e., for $\hat{\Psi}_1$ and $\hat{\Psi}_3$). In contrast, mean bias and MSE are dramatically reduced for the corrected estimators ($\hat{\Psi}_{\text{corr } 1}$ and especially $\hat{\Psi}_{\text{corr } 2}$).

While the MLE for the OR approaches median unbiasedness, the bias-corrected estimators sacrifice that criterion. For example, Table 4 shows that the proportion of $\hat{\Psi}_1$ values below the true OR was 48% for the MLE, vs. 60% and 63% for the ‘corr1’ and ‘corr2’ estimators.

For comparison, the right-most column summarizes the performance of the estimator $\hat{\Psi}_{\text{corr } 2}^*$ with $p=0.55$. This estimator achieves a mean/median bias compromise, with the risk of underestimating the true Ψ_1 approximately equal to the desired threshold of 55%.

Supplementary Figure 2 compares histograms representing the 10,000 standard and bias-corrected OR estimates ($\hat{\Psi}_{ML}$ and $\hat{\Psi}_{corr\ 2}$) based on the simulations summarized in Table 4. The MLE histogram displays a longer and heavier tail, yielding an empirical mean of 7.56. In contrast, the mean bias-corrected estimate was 5.80, again nearly unbiased.

4. DISCUSSION

Ratio estimates of effect are the standard for reporting and interpreting epidemiologic results, reflecting their intuitive appeal as well as their proportionality to relative caseload or risk when the outcome is not common. While we focused primarily on estimates from logistic regression, the basic bias correction [$\hat{\Psi}_{corr}$; eqn. (4)] is applicable in any case of exponentiation of an approximately unbiased estimate. Initial log-scale bias correction and bootstrap variance estimation yields further bias removal on the exponentiated scale. Nonetheless, as noted by a reviewer, an alternative coefficient estimator is necessary if the ML estimate is infinite. The simulations in Tables 3 and 4 yielded no separation problems and thus all could use Cordeiro and McCullagh's (1991) correction. Firth's (1993) approach is now commercially available for logistic and Cox regression (SAS Institute, Inc., 2008), making it attractive for the log-scale correction step when separation is encountered (e.g., Heinze, 1999; Heinze and Schemper, 2002).

Unlike the bias-corrected estimators $\hat{\Psi}_{corr1}$ and $\hat{\Psi}_{corr\ 2}$, $\hat{\Psi}_{ML}$ converges rapidly to median unbiasedness and is transformation invariant (e.g., $1/\hat{\Psi}_{ML}$ is the MLE for $1/\Psi$). But $\hat{\Psi}_{ML}$ is subject to potentially extreme positive mean bias when σ_j^2 is large, and always has higher variance and MSE than the proposed bias-corrected estimators. Thus the bias corrections discussed here will be valuable whenever loss is more proportional to Ψ rather than its log, as we think holds in most policy and planning settings. More generally, eqn. (3) reflects how the positive bias in $\hat{\Psi}_{ML}$ sacrifices traditional mean unbiasedness in favor of median unbiasedness. Arguably, the latter criterion ignores the magnitude of extreme estimates in the sampling distribution, producing an incomplete and perhaps misleading performance measure for an estimator whose realized value may be subject to interpretation with a view toward policy.

When standard errors are large, eqn. (5) indicates that the proposed bias-corrected estimators may entail substantial underestimation risk (median bias) in order to achieve approximate mean unbiasedness. The estimator $\Phi(\hat{\sigma}_j/2)$ (Section 2.2) provides a convenient way to assess this risk and leads to the class of estimators $\hat{\Psi}_{corr}^*$ in (7), which can yield worthwhile reductions in mean bias, variance, and MSE without severe risk of underestimation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was partially supported by National Institute of Environmental Health Sciences Grant 2R01-ES012458-5, an RC4 grant through the National Institute of Nursing Research (1RC4NR012527-01), National Institutes of Health Grant R01-MH079448-01, and a PHS grant (UL 1 RR025008) from the Clinical and Translational Science Award Program, National Institutes of Health, Center for Research Resources. Views expressed are those of the authors.

References

Agresti, A. Categorical Data Analysis. 2. Hoboken, NJ: Wiley; 2002.

- Allison, PD. Convergence Problems in Logistic Regression. In: Altman, M.; Gill, J.; McDonald, MP., editors. Numerical Issues in Statistical Computing for the Social Scientist. Hoboken, NJ: Wiley; 2004. p. 238-252.
- Anderson JA, Richardson SC. Logistic Discrimination and Bias Correction in Maximum Likelihood Estimation. *Technometrics*. 1979; 21:71–78.
- Byth K, McLachlan GJ. The Biases Associated with Maximum Likelihood Methods of Estimation of the Multivariate Logistic Risk Function. *Communications in Statistics A*. 1978; 7:877–890.
- Cook RD, Tsai CL, Wei BC. Bias in Nonlinear Regression. *Biometrika*. 1986; 73:615–623.
- Cordeiro GM, McCullagh P. Bias Correction in Generalized Linear Models. *Journal of the Royal Statistical Society Series B*. 1991; 53:629–643.
- Cox, DR.; Oakes, D. Analysis of Survival Data. London: Chapman & Hall; 1984.
- Davidian, M.; Giltinan, DM. Nonlinear Models for Repeated Measurement Data. New York: Chapman & Hall; 1995.
- Diggle, PJ.; Heagerty, P.; Liang, K-Y.; Zeger, SL. Analysis of Longitudinal Data, Second Edition. New York: Oxford; 2002.
- Efron, B.; Tibshirani, RJ. An Introduction to the Bootstrap. New York: Chapman & Hall; 1993.
- Firth D. Bias Reduction of Maximum Likelihood Estimates. *Biometrika*. 1993; 80:27–38.
- Greenland S. Small-Sample Bias and Corrections for Conditional Maximum- Likelihood Odds-Ratio Estimators. *Biostatistics*. 2000; 1:113–122. [PubMed: 12933529]
- Heinze, G. Technical Report 10/1999. Section for Clinical Biometrics, CeMSIIS, Medical University of Vienna; 1999. The Application of Firth's Procedure to Cox and Logistic Regression.
- Heinze G, Schemper M. A Solution to the Problem of Separation in Logistic Regression. *Statistics in Medicine*. 2002; 21:2409–2419. [PubMed: 12210625]
- Hosmer, DW.; Lemeshow, S. Applied Logistic Regression. New York: Wiley; 2000.
- Jensen JLWV. Sur les Fonctions Convexes et les Inégalités Entre les Valeurs Moyennes. *Acta Mathematica*. 1906; 30:175–193.
- Jewell NP. Small-Sample Bias of Point Estimators of the Odds Ratio from Matched Sets. *Biometrics*. 1984; 40:421–435. [PubMed: 6487726]
- Jewell NP. On the Bias of Commonly Used Measures of Association for 2x2 Tables. *Biometrics*. 1986; 42:351–358.
- King G, Zeng L. Logistic Regression in Rare Events Data. *Political Analysis*. 2001; 9:137–163.
- Lyles RH, Guo Y, Hill AN. A Fresh Look at the Discriminant Function Approach for Estimating Crude or Adjusted Odds Ratios. *The American Statistician*. 2009; 63:320–327.
- McLachlan GJ. A Note on Bias Correction in Maximum Likelihood Estimation with Logistic Discrimination. *Technometrics*. 1980; 22:621–627.
- McCullagh, P.; Nelder, JA. Generalized Linear Models. 2. New York: Chapman & Hall; 1989.
- Read, CB. Median Unbiased Estimators. In: Kotz, S.; Johnson, NL., editors. Encyclopedia of Statistical Sciences. Vol. 5. New York: Wiley; 1985. p. 424-426.
- Rosner, B. Fundamentals of Biostatistics. 5. Pacific Grove, CA: Duxbury; 2000.
- SAS Institute, Inc. SAS/STAT 9.2 User's Guide. SAS Institute; Cary, NC: 2008.
- Schaefer RL. Bias Correction in Maximum-likelihood Logistic Regression. *Statistics in Medicine*. 1983; 2:71–78. [PubMed: 6648121]

Table 1

Cell Counts and Cell Probabilities for Unpaired and Paired Data

Unpaired Data			Paired Data		
	X=1	X=0		Controls (Y=0)	
			Cases (Y=1)	X=1	X=0
Y=1	A (p_{11})	B (p_{10})	X=1	A (p_{11})	B (p_{10})
Y=0	C (p_{01})	D (p_{00})	X=0	C (p_{01})	D (p_{00})

Table 2

$O(n^{-1})$ (second-order) Bias Terms for MLEs and Bias-Corrected Estimators in Unpaired and Paired Data Settings

	Unpaired Data	Paired Data
$\hat{\beta}_{ML}$	$(1/2n)(1/p_{01} + 1/p_{10} - 1/p_{00} - 1/p_{11})$	$(1/2n)(1/p_{01} - 1/p_{10})$
$\hat{\Psi}_{ML}$	$(\psi/n)(1/p_{01} + 1/p_{10})$	$(\psi/n)(1/p_{01})$
$\hat{\Psi}_{corr 1}$	$(\psi/2n)(1/p_{01} + 1/p_{10} - 1/p_{11} - 1/p_{00})$	$(\psi/2n)(1/p_{01} - 1/p_{10})$
$\hat{\Psi}_{corr 2}$	0	0

Table 3
Simulation Results Comparing Estimators for Regression Coefficients and Corresponding Standard Errors

Variable	Regression Coeff. Estimates		Mean of Standard Error Estimates		Mean Width of 95% CI for Ψ (Coverage)	
	$\hat{\beta}_{ML}$ Mean (SD)	$\hat{\beta}_{corr}$ Mean (SD)	ML	Parametric Bootstrap	ML	$\hat{\beta}_{corr}$ w/Parametric Bootstrap
X_1 ($\beta_1=1.75$, $\Psi_1=5.75$)	1.803 (0.646)	1.748 (0.627)	0.636	0.626	25.55 (95.3%)	22.13 (94.8%)
X_2 ($\beta_2=1$, $\Psi_2=2.72$)	1.030 (0.349)	1.001 (0.338)	0.345	0.340	4.44 (95.2%)	4.13 (94.9%)
X_3 ($\beta_3=-.5$, $\Psi_3=0.61$)	-0.516 (1.067)	-0.503 (1.040)	1.051	1.037	8.29 (94.8%)	7.94 (94.6%)

* 10,000 replications with n=200 in each case; Covariate distributions described in text

Table 4

Simulation Results Comparing Performance of Proposed OR Estimators in Logistic Regression *

Variable	Mean Bias [Median Bias] (SD) {Empirical MSE} Percentage of estimates < true value			
	$\hat{\Psi}_{ML}^{\dagger}$	$\hat{\Psi}_{corr1}^{\ddagger}$	$\hat{\Psi}_{corr2}^{\P}$	$\hat{\Psi}_{corr2}^{*//}$
$X_1 (\beta_1=1.75, \Psi_1=5.75)$	1.81 [0.22] (6.11) {40.58} 48%	0.33 [-0.86] (4.64) {21.59} 60%	0.05 [-1.11] (4.48) {20.06} 63%	0.77 [-0.52] (5.04) {26.01} 57%
$X_2 (\beta_2=1, \Psi_2=2.72)$	0.26 [0.05] (1.13) {1.34} 48%	0.09 [-0.11] (1.04) {1.08} 55%	0.00 [-0.18] (0.99) {0.98} 58%	0.05 [-0.15] (1.01) {1.01} 56%
$X_3 (\beta_3=-.5, \Psi_3=0.61)$	0.46 [-0.01] (1.67) {2.99} 50%	0.00 [-0.26] (0.94) {0.88} 70%	0.00 [-0.25] (0.91) {0.83} 70%	0.31 [-0.07] (1.37) {1.98} 55%

* Based on 10,000 replications with n=200 in each case; Covariate distributions described in text

 † Usual MLE for adjusted OR ‡ Bias-corrected estimate computed using $\hat{\beta}_{ML}$ and its standard error [eqn. (4)] ¶ Bias-corrected estimate computed using $\hat{\beta}_{corr}$ and parametric bootstrap-based standard error $^{*//}$ Bias-reduced estimate [eqn. (7)] using $\hat{\beta}_{corr}$ and parametric bootstrap-based standard error; using p=0.55 to limit risk of underestimating true OR to approximately 55% or less