Published in final edited form as:

Stat Methods Med Res. 2017 October ; 26(5): 2257–2269. doi:10.1177/0962280215596550.

Bias corrected estimates for logistic regression models for complex surveys with application to the United States' Nationwide Inpatient Sample

Kevin A. Rader,

Harvard School of Public Health, Boston, MA, U.S.A

Stuart R. Lipsitz,

Brigham and Women's Hospital, Boston, MA, U.S.A

Garrett M. Fitzmaurice,

Harvard Medical School, Boston, MA, U.S.A

David P. Harrington,

Harvard School of Public Health, Boston, MA, U.S.A

Michael Parzen, and

Statistics Department, Harvard University, Cambridge, MA, U.S.A

Debajyoti Sinha

Florida State University, Tallahassee, FL, U.S.A

Abstract

For complex surveys with a binary outcome, logistic regression is widely used to model the outcome as a function of covariates. Complex survey sampling designs are typically stratified cluster samples, but consistent and asymptotically unbiased estimates of the logistic regression parameters can be obtained using weighted estimating equations (WEE) under the naive assumption that subjects within a cluster are independent. Despite the relatively large samples typical of many complex surveys, with rare outcomes, many interaction terms, or analysis of subgroups, the logistic regression parameters estimates from WEE can be markedly biased, just as with independent samples. In this paper, we propose bias-corrected weighted estimating equations for complex survey data. The proposed method is motivated by a study of post-operative complications in laparoscopic cystectomy, using data from the 2009 United States' Nationwide Inpatient Sample (NIS) complex survey of hospitals.

Keywords

Binary responses; Bladder cancer; Population survey; Stratified cluster sampling; Weighted estimating equations

1 Introduction

Complex health surveys are widely used to study the incidence and prevalence of disease, to assess the occurrence of healthy and unhealthy behaviors, and to determine the cost, quality

and utilization of health services. With their large sample sizes, complex health surveys can assess relatively small but clinically important associations among health-related factors, in addition to studying rare events (e.g., orphan diseases) and relatively small subpopulations of interest (e.g., minority populations). Results from these surveys continue to have an important impact on both health care and policy. For example, the National Health and Nutrition Examination Survey (NHANES) is a program of house-hold interview surveys designed to assess the health and nutritional status of adults and children in the United States¹. Data from NHANES have been harnessed to yield estimates of the prevalence of all major diseases, to identify nutritional and other risk factors associated with diseases, and to provide the basis for national standards for measurements of height, weight, and blood pressure. Survey data on height and weight from NHANES have been used to create the growth charts used nationally to monitor children's development. These data have also led to more widespread recognition of the high overweight prevalence, as well as providing the means to track trends in obesity over time. Blood lead data from NHANES were instrumental in developing policy to eliminate lead from gasoline (petrol). Finally, nutrition data from NHANES is used to inform government agencies about the need to change micronutrient (e.g., vitamin and mineral) fortification regulations for the food supply.

In many studies using complex health survey data the response variable of interest is binary. For example, it may be of interest to determine whether or not a patient contracts a disease or complication based on a measurable set of predictors, e.g., age, gender, or environmental exposure factors. In that setting, the logistic regression model is the most commonly used model for predicting a binary outcome from a set of measurable covariates. The focus of this paper is on methods for reducing the bias of the logistic regression model parameter estimates when the data arise from complex surveys.

For independent observations, maximum likelihood is the method of choice for estimating the logistic regression model parameters. However, for independent observations, when the sample size is relatively small or when the binary outcome is either rare or very prevalent (even in large samples), maximum likelihood can yield biased estimates of the logistic regression parameters. In certain cases, when the data has complete or quasi-complete separation, the likelihood may not have a unique solution². Techniques have been developed to remove the first-order term in the Taylor series expansion of the asymptotic bias of the maximum likelihood estimator^{3–5}. This approach is straightforward to implement when observations are sampled independently. For the case of logistic regression with independent subjects, there have been numerous methods proposed for handling these issues, such as exact logistic regression or the bias-correcting approach discussed above⁴; however, such approaches have not been well-studied for binary data from complex sampling schemes.

The focus of this paper is on bias-corrected estimates of the parameters for the logistic regression model when the data arise from complex surveys with stratified and clustered designs. Consistent and asymptotically unbiased estimates of the logistic regression parameters can be obtained using weighted estimating equations (WEE), which incorporate the complex survey weights but naively assume independence among observations within a cluster^{6,7}. A consistent variance estimate can be obtained using a robust sandwich variance estimate that accounts for the stratification, clustering, and weighting^{6,7}. Although

consistent, even with the large samples typical of many complex surveys, with rare outcomes, many interaction terms, or analysis focusing on subgroups, the logistic regression parameters estimates from WEE can be severely biased, just as with independent samples.

Our proposed method is motivated by a study from the 2009 National Inpatient Sample (NIS) that investigated laparoscopic cystectomies to treat bladder cancer⁸; here we use more recent (2010) NIS bladder cancer data. Subjects were identified from the US Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample (NIS), sponsored by the Agency for Healthcare Research and Quality⁹. It is the largest all-payer inpatient care observational cohort in the United States and is representative of approximately 90% of all hospital discharges. The NIS is a 20% stratified probability sample that encompasses approximately 8 million acute hospital stays per year from approximately 1000 hospitals (clusters) in 45 states. In the NIS, there are 60 strata based on five key hospital characteristics. A key hospital characteristic is whether it is a teaching or non-teaching institution. Because approximately 20% of the universe of hospitals are sampled, the weight (or inverse probability of being sampled) for each patient is usually close to five.

In this paper, we analyze patients from the 2010 NIS who received laparoscopic cystectomies to treat bladder cancer ($n = 1263$) in a teaching or non-teaching institution. The primary objective of the study was to compare robot-assisted laparoscopic radical cystectomy (RARC) and open radical cystectomy (ORC) for treatment of bladder cancer within teaching and non-teaching hospitals. We focus on the primary binary endpoint of whether or not the patient contracted a wound infection (yes or no) after surgery. The target of inference is the relative odds of a patient experiencing an infection of the wound area when comparing RARC to ORC. There are three a priori potential confounding factors associated with wound infection, age, sex, and whether the subject had one or more comorbidities, which are summarized for the two groups in the teaching and non-teaching hospitals in Table 1. In our sample from the teaching hospitals there were 54 (6.2%) wound complications in the 878 patients who received standard ORC; there were 6 (4.3%) in the 139 patients that received robot-assisted treatment (RARC). This corresponds to an approximate 30% reduction in the odds (or risk) of infection. In our sample from the non-teaching hospitals there were 21 (9.5%) wound complications in the 221 patients who received standard ORC; there were none (0%) in the 25 patients that received robot-assisted treatment (RARC). The latter result leads to the classic issue of “separation” in the response for two treatment groups, and motivated us to explore a new analytic approach to handle this issue in the complex survey setting.

In Section 2, we briefly describe the complex sampling design, the standard weighted estimating equations (WEE) for the logistic regression model for complex surveys and our bias-corrected version of WEE. In Section 3, we apply this approach to logistic regression analyses of the data from the study of post-operative complications in laparoscopic cystectomy⁸. In Section 4, we present results of a small-scale simulation study of our bias correction for the logistic regression model. In the example and simulations, we compare our approach to the standard WEE for complex surveys without bias correction.

2 Methods

The most common type of complex survey design is a stratified cluster design. Further, more complex multi-stage designs can be approximated as a stratified cluster design¹⁰. Thus here we use notation for stratified cluster designs. We let y_{hij} represent the Bernoulli outcome for the j^{th} subject, ($j = 1, \dots, m_{hi}$), in the i^{th} cluster, ($i = 1, \dots, n_h$), within the h^{th} stratum, ($h = 1, \dots, H$). Note that we assume there are H strata, n_h clusters in stratum h , and m_{hi} subjects in cluster i of stratum h . Let the indicator variable δ_{hij} equal 1 if subject hij is selected into the sample and equal 0 otherwise. The probability of being selected into the survey, $P(\delta_{hij} = 1) = p_{hij}$, is fixed by the study design and may depend on the outcome of interest, the covariates, or additional variables (screening variables, for example) not in the logistic regression model for the outcome of interest. Thus, each subject in the sample has a known ‘weight’ $w_{hij} = \delta_{hij}/p_{hij}$. We let π_{hij} be the probability that $Y_{hij} = 1$, which follows the standard logistic regression model:

$$\pi_{hij} = P(Y_{hij} = 1 | \mathbf{x}_{hij}, \beta) = \frac{\exp(\beta' \mathbf{x}_{hij})}{1 + \exp(\beta' \mathbf{x}_{hij})}$$

where $\mathbf{x}_{hij}$ is a $(k+1) \times 1$ vector of covariates including the constant term for the hij^{th} observation, and β is a $(k+1) \times 1$ parameter vector including the intercept term. In principle, other link functions could be chosen; however, the focus of this paper is on logistic regression.

To obtain consistent estimates of β in complex surveys, one needs to incorporate the subject-specific sampling weights, w_{hij} , into the logistic regression estimating equations. Weighting estimating equations (WEE), which naively assume subjects are independent, have been shown to give consistent estimates¹¹, and are of the form, $U(\hat{\beta}) = 0$, where:

$$U(\beta) = \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \mathbf{x}_{hij} (y_{hij} - \pi_{hij}). \quad (1)$$

The standard multivariable estimating equations can be modified to correct for the first-order bias³. This can be done by replacing the responses, y_{hij} , with “pseudoresponse”, y_{hij}^* :

$$y_{hij}^* = y_{hij} + a_{hij}$$

where a_{hij} represents an adjustment to the observed response, y_{hij} , and is defined as:

$$a_{hij} = 0.5 \left(\text{tr} \left[\text{Var}(\hat{\beta}) D''_{hij} \right] \right) \quad (2)$$

where D''_{hij} is the second derivative matrix of the logistic function, π_{hij} , with respect to β . Specifically, for the logistic regression model, $D''_{hij} = \mathbf{x}_{hij} \mathbf{x}'_{hij} \pi_{hij} (1 - \pi_{hij}) (1 - 2\pi_{hij})$, and the adjustment factor, $a_{hij} = 0.5 \pi_{hij} (1 - \pi_{hij}) (1 - 2\pi_{hij}) [\text{Var}(\mathbf{x}_{hij} \hat{\beta})]$ (see Appendix 1). Note, in generalized linear model terminology, $\text{logit}(\pi_{hij}) = \mathbf{x}_{hij} \beta$ is referred to as the “linear predictor”. Thus, the adjustment term is a simple function of π_{hij} and the variance of the estimated linear predictor. When there are no sampling weights involved, the adjustment term a_{hij} is equivalent to the adjustment for ordinary logistic regression⁴. Replacing y_{hij} in (1) with y_{hij}^* , the bias-reduced estimating equations become:

$$\begin{aligned} U(\beta)^* &= \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \mathbf{x}_{hij} (y_{hij}^* - \pi_{hij}) \\ &= \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \mathbf{x}_{hij} \{y_i - \pi_{hij} + q_{hij} (0.5 - \pi_{hij})\} \end{aligned} \quad (3)$$

where $q_{hij} = \pi_{hij} (1 - \pi_{hij}) \text{Var}(\mathbf{x}_{hij} \hat{\beta})$ (see Appendix 1 for derivation of (3)). For standard logistic regression, the term q_{hij} reduces to the leverage for observation hij .

Bias-corrected estimates can also be calculated by splitting each of the original observations into two new observations: one taking value y_{hij} and the other taking value $1 - y_{hij}$ but with weights $1 + q_{hij}/2$ and $q_{hij}/2$, respectively¹². Extending these results to complex surveys with weights w_{hij} , we propose the use of the weights $w_{hij}(1 + q_{hij}/2)$ and $w_{hij}(q_{hij}/2)$ for y_{hij} and $1 - y_{hij}$ respectively. Thus each individual contributes $\{(y_{hij} - \pi_{hij}) w_{hij}(1 + q_{hij}/2) + (1 - y_{hij} - \pi_{hij}) w_{hij}(q_{hij}/2)\}$ to the score function, which can be shown to be mathematically equivalent to the following weighted estimating equations:

$$\begin{aligned} U(\beta)^* &= \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \{(y_{hij} - \pi_{hij}) w_{hij}(1 + q_{hij}/2) + (1 - y_{hij} - \pi_{hij}) w_{hij}(q_{hij}/2)\} \mathbf{x}_{hij} \\ &= \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \{(y_{hij} - \pi_{hij}) + q_{hij}(1/2 - \pi_{hij})\} w_{hij} \mathbf{x}_{hij}, \end{aligned} \quad (4)$$

(see Appendix 1 for derivation of (4)).

Even though complex surveys commonly have large sample sizes, the issue of separation can arise when domains are small or subgroup analyses are performed. The bladder cancer study mentioned in the Introduction is an example where this has occurred, as the number of wound infections in the robotic treatment arm is zero. In the bias-corrected WEE in (4), $y_{hij} = 1$ and $y_{hij} = 0$ have positive weights, which is equivalent to there being a non-zero number of successes ($y_{hij} = 1$) and failures ($y_{hij} = 0$) at each value of x_{hij} . Because of this property, the adjusted weighted estimating equations (equivalent to standard logistic regression with weights) we propose have a unique, finite solution (assuming the design matrix is full rank)¹³. By splitting each observation into two, this eliminates the problem of separation when the response variable is composed of all successes or all failures for a specific

combination of the covariates, and allows for the use of standard complex survey software, e.g., `svyglm` in *R*, to perform the analysis.

Theory suggests that using a consistent estimate of the true $\text{Var}(\mathbf{x}_{hij} \hat{\beta})$ in q_{hij} should reduce the bias³. Two approaches for consistently estimating $\text{Var}(\hat{\beta})$ in q_{hij} and thus $\text{Var}(\mathbf{x}_{hij} \hat{\beta})$ in q_{hij} are the standard sandwich estimator based on a Taylor series and the small-sample bias-corrected estimator of variance developed by Morel et al.^{6,14}. Because the focus of this paper is on bias-reduction, and the standard sandwich variance estimator can be highly variable for rare events, or a small number of large clusters, we use the small-sample bias-corrected variance estimator proposed by Morel et al. to estimate $\text{Var}(\hat{\beta})$ instead of the standard sandwich variance estimator¹⁴. In particular, in simulations the small-sample bias-corrected variance estimator has been shown to be much less variable and more stable than the standard sandwich variance estimator¹⁴. Thus, to calculate q_{hij} we consider two approaches: (a) naively assuming independence among the observations and (b) using the bias-corrected sandwich variance estimator¹⁴. A third approach is to use the ad hoc bias-

reduction method proposed by Clogg et al.¹⁵, setting $q_{hij} = p_0 k / (\sum_{h=1}^H \sum_{i=1}^{n_h} m_{hi})$ where p_0 is an estimate or prespecified value for the overall proportion of successes. Finally, an algorithm for the bias-corrected estimators is outlined in Appendix 2.

3 Application to Bladder Cancer Study

In this section, we apply the proposed methods to the analysis of the radical cystectomy data from the National Inpatient Sample (NIS) described in the Introduction. This analysis of the NIS includes 1263 patients (using the weights, representing 6505 patients in the population) undergoing radical cystectomy to treat bladder cancer throughout the United States. The outcome of interest is binary: whether or not the patient experienced wound infections post-surgery (1=infection, 0 = no infection) while staying at the hospital. Our main comparison of interest is to determine whether the relative odds of wound infections differs between the two types of cystectomy: standard open radical cystectomy (ORS) and robot-assisted radical cystectomy (RARC). Based on an earlier study, a priori, we conjectured that robot-assisted surgeries would have a lower rate of infection⁸. However, the potential benefits of RARC may be somewhat different at teaching versus non-teaching hospitals. As a result, we analyzed the data separately for the two types of institutions. Based on the earlier study, the covariates most predictive of wound infection are surgery type (ORS and RARC), age, sex, and comorbidity⁷.

To examine the relationship between post-surgery wound infection and these 4 covariates, we fit the logistic regression model,

$$\text{logit}[\pi_i] = \text{logit}\{P[Y_i=1|\mathbf{x}_i, \beta]\} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i}, \quad (5)$$

where x_{1i} is the surgery type ($x_{1i} = 1$ for robot-assisted and 0 for standard open radical cystectomy), x_{2i} is the age of the patient in years, x_{3i} is an indicator of the sex of the i^{th} patient ($x_{3i} = 1$ for females and 0 for males), and x_{4i} is an indicator that the i^{th} patient had

one or more comorbidities. Estimates of β were obtained using the two bias-corrected approaches we propose, in addition to the Clogg et al. approach with

$q_{hij} = p_0 k / (\sum_{h=1}^H \sum_{i=1}^{n_h} m_{hi})$ and the standard WEE estimates (the latter were obtained using R `svyglm`); analyses were conducted separately for teaching and non-teaching hospitals.

Before presenting the results of these logistic regression analyses, we consider the unadjusted effects of surgery type on post-surgery wound infection. Ignoring the complex sampling design, analysis of the 2×2 contingency table for teaching hospitals yields an estimated odds ratio of 0.69 (exact 95% CI: 0.24, 1.64). Properly accounting for the complex sampling design using our proposed bias-corrected method (with Morel variance) yields an estimated odds ratio of 0.74 (95% CI: 0.25, 2.15). The latter result suggests that the approximate 25% reduction in the odds can be explained due to sampling variability. Similar analysis of the 2×2 contingency table for non-teaching hospitals yields an estimated odds ratio of 0 (exact 95% CI: 0.00, 1.66) because there were no complications in the robotic arm. In contrast, properly accounting for the complex sampling design using our proposed bias-corrected method (with Morel variance) yields an estimated odds ratio of 0.18 (95% CI: 0.10, 0.35). This result suggests a significant 82% decline in the odds of infection with RARC versus ORC in patients treated in non-teaching hospitals.

Next, we consider the results of the logistic regression analyses adjusting for age, sex, and comorbidity. Table 2 gives the estimates of β obtained using the two bias-corrected approaches we propose, the Clogg et al. approach, and the standard WEE estimates; the results are presented separately for teaching and non-teaching hospitals. Of note, there were no convergence problems with the three bias-corrected methods. However, because there were no complications in the robotic arm for the non-teaching hospitals, the coefficient for β_1 was converging to $-\infty$ for WEE; the results for WEE reported for the non-teaching hospitals in Table 2 are the estimates at the 25th iteration (the default maximum number of iterations in R's `svyglm` function in the package `survey`). For all of the methods, the standard errors reported in Table 2 are based on the variance estimator of Morel, et al¹⁴.

Using the independence variance when calculating the adjustment term, the estimated odds ratio (OR) for surgery type in teaching hospitals, controlling for age, sex and comorbidity, is $e^{-0.306} = 0.74$; when the small-sample bias-corrected variance is used the OR is also estimated to be $e^{-0.305} = 0.74$. For the Clogg estimator, the OR is estimated to be $e^{-0.370} = 0.69$. When comparing the estimates of β to their standard errors, the three bias-corrected approaches produce similar estimates, and all lead to the same conclusion that there is no association between surgery type and odds of wound infection. In contrast, the estimated OR for surgery type in non-teaching hospitals, using the independence variance adjustment term, is $e^{-1.759} = 0.17$ (95% CI: 0.09, 0.35); when the small-sample bias-corrected variance is used the OR is estimated to be $e^{-1.711} = 0.18$ (95% CI: 0.10, 0.34). For the Clogg estimator, the OR is estimated to be $e^{-2.490} = 0.08$ (95% CI: 0.05, 0.14). In this case, the independence and small-sample bias-corrected variance adjustment terms yield very similar estimates of effect: a significant 83% and 82% decline, respectively, in the odds of infection with RARC versus ORC in patients treated in non-teaching hospitals (similar to result of unadjusted

analysis). The results for the Clogg method and the standard WEE suggest there is a significant association between surgery type and odds of wound infection; when comparing the estimates of β to their standard errors, we see that the standard WEE (the 25th iteration in R `svyglm`) produces a much more significant result than the bias-corrected approaches. Because the estimated OR is on the boundary, caution must be exercised in making inferences based on the WEE method.

In summary, the results of analyses of the bladder cancer data highlight how the standard WEE approach and the bias-corrected methods can produce somewhat different estimates of effects. On the basis of results from the proposed small-sample bias-corrected variance adjustment, there is no evidence of association between surgery type and odds of wound infection in patients treated in teaching hospitals but a very discernible reduction (82% decline) in the odds of infection with RARC versus ORC in patients treated in non-teaching hospitals. Using combined data on patients from both teaching and non-teaching hospitals, a test of heterogeneity of the effect of surgery type (adjusted for age, sex and comorbidity) yielded $Z = 2.32$, $p < 0.05$. To examine the finite sample bias of these approaches, we conducted a simulation study; the results of the simulation study are reported in the next section.

4 Simulation Study

In this section, we study the empirical relative bias in estimating β using standard logistic regression models incorporating the complex survey structure (WEE) and our bias-reduced approach using 3 different variance estimators when calculating the multiplicative weighting factor, q_{hij} : variance under independence (Independence), variance using the bias-corrected sandwich estimator of Morel et al. (Morel), along with the Clogg et al. approach (Clogg)^{14,15}. For simplicity, in the simulation study, we used a cluster design without stratification and weighting, where sampling of clusters was performed without replacement from a finite population of clusters.

For the simulations, the true marginal logistic model for any subject in the population is

$$\text{logit}(P[Y_{ij}=1|\mathbf{x}_{ij}]) = \beta_0 + \sum_{k=1}^{10} \beta_k x_{ijk}, \quad (6)$$

where the ten x_{ijk} 's are independent Bernoulli($p_x = 0.15$) variables. The intercept β_0 was chosen so that the average $P[Y_{ij}=1]$ equals 0.10, corresponding to an overall prevalence of 10%, and for simplicity we set $\rho_0 = 0.1$ for the Clogg approach to reflect this prevalence. This marginal model is similar to that used in a simulation study performed by Heinze and Schemper¹². For simplicity, we set all ten β_k equal to the same value,

$$\text{logit}(P[Y_{ij}=1|\mathbf{x}_{ij}]) = \beta_0 + \beta \sum_{k=1}^{10} x_{ijk}$$

To simulate the clustered data, we use the random intercept logistic regression model proposed by Wang and Louis and further developed by Parzen et al.^{16,17}. In particular, the conditional subject-specific logistic regression model is

$$\text{logit}(P[Y_{ij}=1|\mathbf{x}_{ij}])=b_i+\left(\beta_0+\beta\sum_{k=1}^{10}x_{ijk}\right)/\phi, \quad (7)$$

where, given the subject-specific random effect b_i , the Y_{ij} 's from the same cluster are independent Bernoulli random variables. When b_i follows a 'bridge' distribution, the marginal logistic regression equals that given in (6)¹⁶. The bridge random variable has mean 0 and ϕ is the rescaling parameter. In particular,

$$\text{Var}(b_i)=\frac{\pi^2}{3}\left(\frac{1}{\phi^2}-1\right),$$

so that the larger the value of ϕ , ($0 < \phi < 1$), the smaller the variance (and the lower the correlation between pairs of random variables in the same cluster). We denote the population number of clusters by N , which we set to $N=400$, the number of sampled clusters by n , and the cluster size by m_i (we assume all clusters are of the same size, and all members of the cluster are sampled).

We conducted 24 simulation configurations varying the following conditions: the effect of the covariates, $\beta_k = \beta = \{\ln(2), \ln(4), \ln(16)\}$ (recall, we set all ten β_k to the same value); cluster sizes, $m_i = \{5, 10\}$; the bridge distribution's scaling parameter, $\phi = \{0.7, 0.9\}$; and the number of clusters sampled, $n = 40$ and $n = 80$. For each simulation configuration, 2000 simulation replications were performed. The convergence criterion for WEE is that the relative change in the log-likelihood between successive iterations is less than 0.000001 (the default maximum number of iterations is 25 in R's `svyglm` function in the package `survey`); we report the percentage of simulation replications in which this convergence criterion was met. In presenting results for the standard WEE, we summarize the results for WEE conditional on convergence.

Tables 3, 4, and 5 present the relative biases for β_2 defined as $100(\hat{\beta}_2 - \beta_2)/\beta_2$, the mean square error of the estimates, and the empirical coverage probabilities of 95% Wald confidence intervals for all the simulation study specifications, respectively. Without loss of generality we report results for β_2 only; any of the β_k could have been selected for bias reporting since the model is symmetric across covariates (all covariates are independent and have the same Bernoulli distribution with all ten $\beta_k = \beta$). The results indicate that the relative bias is greatly reduced, by an order of magnitude, when using the independence and small-sample bias-corrected variance adjustment terms. For both of these bias-reduced approaches the relative bias is less than 2% across most of the simulation configurations (and less than 5% across all configurations). The standard WEE method resulted in the most biased estimates, with relative bias as high as 200% and often greater than 10% for $\beta = \ln(16)$. For

$\beta = \ln(4)$ there are some convergence issues for the WEE method and it has moderate relative bias typically between 5% and 10%. For $\beta = \ln(2)$ there are no convergence issues for the WEE method and thus low relative bias. The Clogg method had relative bias that fell somewhere between the standard WEE approach and the bias-reduced approaches proposed here. Interestingly, this ordering of the amount of bias across these methods agrees with previous simulation results for independent observations as seen in Heinze and Schemper¹². In general, the Clogg method tended to be more biased and have higher MSE than the other bias-reduced approaches. When both the number of clusters and the cluster size were relatively small, the standard WEE had problems with convergence when β_k was far from the null (see Table 5). For example, the standard WEE method converged about 80% of the time when $\phi = 0.7$, $m_i = 5$, $n = 40$, and $\beta = \ln(16)$. Overall, results of our simulations suggest that the bias-reduced approach using either the independent variance estimate or the Morel variance estimate to calculate q_{hij} is the preferred method for performing the analysis.

Although Wald confidence intervals are known to exhibit unreliable and aberrant behavior in logistic regression with large β 's, we found that in most of the simulation configurations the coverage probabilities using the Wald confidence intervals with Morel-adjusted variance agree with the nominal 95% level for the bias-corrected approaches^{12,18,19}. However, we caution against generalizing this finding based on the results of a single simulation study. One potential alternative to Wald confidence intervals is the so-called pseudo-score confidence intervals developed by Agresti and Ryu²⁰. Pseudo-score confidence intervals can be obtained by inverting a test that uses a quasi-likelihood Pearson chi-squared statistic that compares fitted values for the working model with fitted values of the model when a parameter of interest takes various fixed values. The performance of pseudo-score confidence intervals in the complex survey setting is a topic of future research.

Finally, although the focus of this paper, and the simulation study, is on bias-reduction in estimation of logistic regression parameters for binary data in the complex survey setting, we also considered the bias in the estimated probabilities. Specifically, using a similar simulation design configuration as described earlier, we considered the bias in the estimated probabilities when the ten x_{ijk} 's were fixed at the following set of values: $x_{ij1} = 1, \dots, x_{ij4} = 1$, $x_{ij5} = 0, \dots, x_{ij10} = 0$ (4 of the x 's set to 1, all others to zero). The intercept β_0 was chosen so that the average $P[Y_{ij} = 1]$ approximately equals 0.3 and 0.8 when $\beta_k = \beta = \{\ln(2), \ln(4)\}$ respectively. The results of these simulations (not reported) suggest that the bias in the estimated probabilities is more comparable between the standard WEE and the bias-reduced approach using either the independent variance estimate or the Morel variance estimate. That is, although the standard WEE yielded discernibly greater bias in estimating the β 's, this did not translate into discernibly greater bias in the estimated probabilities.

5 Discussion

In this paper we have described a simple implementation of bias correction in the logistic regression model for complex surveys. By incorporating an adjustment term to the weighted estimating equations, we derived a bias correction based on univariate Bernoulli distributions. This bias correction splits each of the original observations into two new observations: the original response, y_i with the original sampling weight, w_{hi} , times a

multiplicative factor, $1+q_{hij}/2$, and a pseudo-response, $1-y_{hij}$ with the original sampling weight times the multiplicative factor minus one, or $q_{hij}/2$. Since both the response and pseudo-response have weights that are guaranteed to be positive, the problematic issue of separation is eliminated¹³. These pseudo-responses and weights are relatively easy to calculate, and this approach leads to an iterative algorithm that is straightforward to implement. Because WEE is the most widely used estimation approach for logistic regression models in complex surveys, the approach to correct for bias described in this paper should be useful in applications where there are rare outcomes, many interaction terms, or the focus of analysis is on particular subgroups of interest. We note that there is an alternative approach based on the large class of Bayesian methods that yield design-consistent estimates of the logistic regression model parameters²¹. These methods require some care in the selection of prior distribution and are most appealing when a relatively weak prior distribution is chosen.

Although not specifically discussed in this paper, the proposed method can also be used for any regression model for binary outcomes in complex surveys, including models that adopt non-canonical link functions, such as probit or complementary log-log links. Kosmidis described bias-corrected estimating equations for non-canonical links for binary data with independent observations, where the original observations are split into y_i and $1-y_{hij}$ with weights for each that are a function of a_{hij} in (2)²². The proposed approach could be used to extend these results by incorporating the complex survey sampling weights into the weights for y_i and $1-y_{hij}$. Further, the approach can be extended to other generalized linear models using weighted estimating equations for complex surveys. The bias-corrected approach for complex surveys would be similar to that given in Kosmidis and Firth for other generalized linear models based on specific link functions²³. This can be implemented by creating a pseudo-response (a function of the outcome and a_{hij}) to correct for the first-order bias.

Finally, the results of the simulations demonstrate that the proposed method can reduce the finite sample bias of WEE for estimating logistic regression parameters for binary data in the complex survey setting. WEE estimates can be biased due to the issue of separation or quasi-separation, a problem that can occur in large complex surveys when the outcomes are rare or subgroup analyses are performed. The bias-corrected methods perform discernibly better than the standard WEE approach for binary data, suggesting that they could be adopted as the method of choice in regression analyses of binary outcomes in complex surveys. Because of its computational simplicity, the standard bias-corrected logistic regression approach of Firth, where q_{hij} is estimated under the naive assumption of independence, may be somewhat easier to implement using standard statistical software for logistic regression with complex survey data⁴.

Acknowledgments

We thank Edward Giovannucci and Caprice Greenberg for advice on the analysis of the NIS bladder cancer data. We are grateful for the support provided by grant CA 160679 from the U.S. National Institutes of Health.

References

1. Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National Health and Nutrition Examination Survey: Sample design, 2011–2014. National Center for Health Statistics. Vital and Health Statistics Series. 2014; 2(168)
2. Albert A, Anderson JA. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*. 1984; 71:1–10.
3. Box MJ. Bias in nonlinear estimation. *Journal of the Royal Statistical Society Series B (Methodology)*. 1971; 33:171–201.
4. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993; 80:27–38.
5. Kosmidis I, Firth D. Bias reduction in exponential family nonlinear models. *Biometrika*. 2009; 96:793–804.
6. Binder D. On the variance of asymptotically normal estimators from complex surveys. *International Statistical Review*. 1983; 51:279–292.
7. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22.
8. Yu H, et al. Comparative analysis of outcomes and costs following open radical cystectomy versus robot-assisted laparoscopic radical cystectomy: results from the US Nationwide Inpatient Sample. *European Urology*. 2012; 61:1239–1244. [PubMed: 22482778]
9. HCUP, NIS. Database Documentation. Healthcare Cost and Utilization Project (HCUP). SID Database Documentation. Rockville, MD: Agency for Healthcare Research and Quality; 2007.
10. Kish, L. Survey Sampling. New York, NY: Wiley; 1965.
11. Shah, BV., Barnwell, BG., Bieler, GS. SUDAAN, Software for the Statistical Analysis of Correlated Data: User's Manual, Release 7.0. Research Triangle Institute; 1996.
12. Heinze G, Schemper M. Comparing the importance of prognostic factors in Cox and logistic regression using SAS. *Computer Methods and Programs in Biomedicine*. 2003; 71:155–163. [PubMed: 12758137]
13. Wedderburn RWM. On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika*. 1976; 63:27–32.
14. Morel JG, Bokossa MC, Neerchal NK. Small sample correction for the variance of GEE estimators. *Biometrical Journal*. 2003; 45:395–409.
15. Clogg CC, Rubin DB, Schenker N, Schultz B, Weidman L. Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression. *Journal of the American Statistical Association*. 1991; 86:68–78.
16. Wang Z, Louis TA. Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. *Biometrika*. 2003; 90:765–775.
17. Parzen M, et al. A generalized linear mixed model for longitudinal binary data with a marginal logit link function. *The Annals of Applied Statistics*. 2011; 5:449–467. [PubMed: 21532998]
18. Hauck WW, Donner A. Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association*. 1977; 72:851–853.
19. Bull SB, Lewinger JP, Lee S. Confidence intervals for multinomial logistic regression in sparse data. *Statistics in Medicine*. 2007; 26:903–918. [PubMed: 16489602]
20. Agresti A, Ryu E. Pseudo-score inference for parameters in discrete statistical models. *Biometrika*. 2010; 97:215–222.
21. Little RJ. Calibrated Bayes, an alternative inferential paradigm for official statistics. *Journal of Official Statistics*. 2012; 28:309–334.
22. Kosmidis I. On iterative adjustment of responses for the reduction of bias in binary regression models. Technical Report 09–36. CRISM working paper series. 2009
23. Kosmidis I, Firth D. Multinomial logit bias reduction via the Poisson log-linear model. *Biometrika*. 2011; 98:755–759.

Appendix 1

Second derivative matrix of π_{hij} with respect to β and derivation of weighted estimating equations

For the logistic regression model, D''_{hij} reduces to

$$\begin{aligned}
 D''_{hij} &= \frac{\partial^2}{\partial \beta^2} [\pi_{hij}] = \frac{\partial^2}{\partial \beta^2} \left([1 + \exp(-\mathbf{x}_{hij}\beta)]^{-1} \right) \\
 &= \frac{\partial}{\partial \beta} \left(\mathbf{x}_{hij} [1 + \exp(-\mathbf{x}_{hij}\beta)]^{-1} \exp(-\mathbf{x}_{hij}\beta) \right) \\
 &= \frac{\partial}{\partial \beta} (\mathbf{x}_{hij} \pi_{hij} [1 - \pi_{hij}]) \\
 &= \frac{\partial \pi_{hij}}{\partial \beta} \frac{\partial}{\partial \pi_{hij}} (\mathbf{x}_{hij} \pi_{hij} [1 - \pi_{hij}]) \\
 &= (\mathbf{x}_{hij} \pi_{hij} [1 - \pi_{hij}]) ([1 - 2\pi_{hij}] \mathbf{x}'_{hij}) \\
 &= \mathbf{x}_{hij} \mathbf{x}'_{hij} \pi_{hij} (1 - \pi_{hij}) (1 - 2\pi_{hij}),
 \end{aligned}$$

and the adjustment factor, a_{hij} , simplifies to:

$$\begin{aligned}
 a_{hij} &= 0.5 \left(\text{tr} \left[\text{Var}(\hat{\beta}) D''_{hij} \right] \right) \\
 &= 0.5 \left(\text{tr} \left[\text{Var}(\hat{\beta}) \mathbf{x}_{hij} \mathbf{x}'_{hij} \pi_{hij} (1 - \pi_{hij}) (1 - 2\pi_{hij}) \right] \right) \\
 &= 0.5 \pi_{hij} (1 - \pi_{hij}) (1 - 2\pi_{hij}) \left[\text{Var}(\mathbf{x}_{hij} \hat{\beta}) \right],
 \end{aligned}$$

since $\text{Var}(\mathbf{x}_{hij} \hat{\beta})$ is a scalar.

Replacing y_{hij} in (1) with y_{hij}^* , the bias-reduced estimating equations become:

$$\begin{aligned}
 U(\beta)^* &= \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \mathbf{x}_{hij} \left(y_{hij}^* - \pi_{hij} \right) \\
 &= \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \mathbf{x}_{hij} (y_{hij} + a_{hij} - \pi_{hij}) \\
 &= \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \mathbf{x}_{hij} \left\{ y_{hij} + 0.5 \pi_{hij} (1 - \pi_{hij}) (1 - 2\pi_{hij}) \left(\text{Var}(\mathbf{x}_{hij} \hat{\beta}) \right) - \pi_{hij} \right\} \\
 &= \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \mathbf{x}_{hij} \left\{ y_{hij} - \pi_{hij} + \left[\pi_{hij} (1 - \pi_{hij}) \left(\text{Var}(\mathbf{x}_{hij} \hat{\beta}) \right) \right] (0.5 - \pi_{hij}) \right\} \\
 &= \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \mathbf{x}_{hij} \{ y_{hi} - \pi_{hij} + q_{hij} (0.5 - \pi_{hij}) \}
 \end{aligned}$$

where $q_{hij} = \pi_{hij} (1 - \pi_{hij}) \text{Var}(\mathbf{x}_{hij} \hat{\beta})$.

Finally, splitting each of the original observations into two new observations, each individual contributes $\{(y_{hij} - \pi_{hij})w_{hij}(1 + q_{hij}/2) + (1 - y_{hij} - \pi_{hij})w_{hij}(q_{hij}/2)\}$ to the score function, which can be shown to be mathematically equivalent to the following weighted estimating equations:

$$\begin{aligned} U(\beta)^* &= \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \{(y_{hij} - \pi_{hij})w_{hij}(1 + q_{hij}/2) + (1 - y_{hij} - \pi_{hij})w_{hij}(q_{hij}/2)\} \mathbf{x}_{hij} \\ &= \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \{y_{hij} - \pi_{hij} - \pi_{hij}q_{hij}/2 + q_{hij}/2 - \pi_{hij}q_{hij}/2\} w_{hij} \mathbf{x}_{hij} \\ &= \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \{y_{hij} - \pi_{hij} - \pi_{hij}q_{hij} + q_{hij}/2\} w_{hij} \mathbf{x}_{hij} \\ &= \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \{(y_{hij} - \pi_{hij}) + q_{hij}(1/2 - \pi_{hij})\} w_{hij} \mathbf{x}_{hij}. \end{aligned}$$

Appendix 2

Algorithm for Bias-Corrected Estimates

To obtain the first-order bias-corrected estimates of β , one can iterate between updating q_{hij} given a current estimate of β and $\widehat{\text{Var}}(\mathbf{x}_{hij}\hat{\beta})$, and then re-estimating β and $\widehat{\text{Var}}(\mathbf{x}_{hij}\hat{\beta})$ given the updated q_{hij} by solving (4), until the estimates of β converge.

In particular, we start by initializing $q_{hij} = k / (\sum_{h=1}^H \sum_{i=1}^{n_h} m_{hi})$, which is the average value of the q_{hij} when the observations are independent. This initial estimate is equivalent to the ad hoc bias-reduction approach proposed by Clogg et al. with independent observations¹⁵. Note

that $\sum_{h=1}^H \sum_{i=1}^{n_h} m_{hi}$ is the total sample size. We then iterate between the following two steps until convergence of $\hat{\beta}$ is obtained:

1. Calculate the complex survey based estimates of β , but with modified survey weights of $w_{hij}(1 + q_{hij}/2)$ for the original y_{hij} and weights of $w_{hij}(q_{hij}/2)$ for the pseudo-observations $1 - y_{hij}$, where w_{hij} is the original sampling weight (this is straight-forward to implement using `svyglm` in *R* or a similar procedure in another software program).
2. Recalculate q_{hij} based on the estimates, $\hat{\pi}_{hij}$ and $\widehat{\text{Var}}(\mathbf{x}_{hij}\hat{\beta})$, from the logistic regression model estimates in the previous step.

Note that in this iterative procedure, $\text{Var}(\hat{\beta})$ and thus q_{hij} is calculated either under independence or using the Morel et al. bias-corrected sandwich variance estimator¹⁴. Further, after convergence, the Morel et al. sandwich variance estimator is used to estimate the variance of $\hat{\beta}$ in making inferences from both approaches¹⁴. That is, for the setting where our proposed approach is most applicable (rare events, small domains, many interaction terms), the bias-corrected variance estimator of $\hat{\beta}$ proposed by Morel et al. is needed¹⁴.

Table 1

Baseline characteristics (means/percentages and 95% confidence intervals) of bladder cancer patients treated with radical cystectomy in teaching and non-teaching hospitals in the National Inpatient Sample (NIS).

Teaching Hospitals:		
	Open Radical Cystectomy (ORC), <i>n</i> = 878	Robot-assisted Radical Cystectomy (RARC), <i>n</i> = 139
Age, years	67.8 (66.6, 69.1)	68.8 (67.4, 70.3)
Female, %	13.3 (10.9, 16.0)	12.0 (6.6, 21.0)
One or more comorbidities, %	19.8 (16.8, 23.0)	21.1 (16.2, 27.0)
Non-Teaching Hospitals:		
	Open Radical Cystectomy (ORC), <i>n</i> = 221	Robot-assisted Radical Cystectomy (RARC), <i>n</i> = 25
Age, years	69.1 (67.8, 70.5)	71.0 (67.3, 74.8)
Female, %	17.4 (13.4, 22.0)	15.5 (7.0, 31.0)
One or more comorbidities, %	16.1 (11.3, 22.0)	4.2 (0.6, 24.0)

Note: Results are reported as population estimates using survey weights, strata, and cluster variables.

Comparison of the logistic regression parameter estimates for the bladder cancer data for teaching and non-teaching hospitals from the National Inpatient Survey (NIS).

Table 2

Effect	Approach	Estimate	SE	Z-statistic	P-value
Teaching Hospitals:					
Intercept	Standard WEE	-3.024	0.851	-3.555	<0.001
<i>Bias-Reduced</i>	Independent Var.	-2.780	0.960	-2.896	0.005
	Morel Var.	-2.789	0.778	-3.586	<0.001
	Clogg Var.	-3.004	0.839	-3.581	<0.001
Robot	Standard WEE	-0.373	0.589	-0.634	0.529
<i>Bias-Reduced</i>	Independent Var.	-0.306	0.424	-0.720	0.474
	Morel Var.	-0.305	0.539	-0.567	0.573
	Clogg Var.	-0.370	0.583	-0.635	0.528
Age	Standard WEE	-0.017	0.011	-1.553	0.125
<i>Bias-Reduced</i>	Independent Var.	-0.017	0.012	-1.426	0.159
	Morel Var.	-0.017	0.010	-1.634	0.107
	Clogg Var.	-0.017	0.011	-1.553	0.125
Female	Standard WEE	0.399	0.324	1.231	0.223
<i>Bias-Reduced</i>	Independent Var.	0.432	0.340	1.268	0.209
	Morel Var.	0.434	0.306	1.415	0.162
	Clogg Var.	0.397	0.322	1.231	0.223
Comorbid	Standard WEE	1.471	0.456	3.225	0.002
<i>Bias-Reduced</i>	Independent Var.	1.268	0.645	1.966	0.054
	Morel Var.	1.265	0.373	3.395	<0.001
	Clogg Var.	1.449	0.444	3.264	0.002
Non-Teaching Hospitals:					
Intercept	Standard WEE	-3.675	2.073	-1.773	0.081
<i>Bias-Reduced</i>	Independent Var.	-3.346	1.713	-1.953	0.055
	Morel Var.	-3.023	1.727	-1.750	0.084
	Clogg Var.	-3.617	2.110	-1.714	0.091
Robot	Standard WEE	-16.337	0.400	-40.838	<0.001

Effect	Approach	Estimate	SE	Z-statistic	P-value
<i>Bias-Reduced</i>	Independent Var.	-1.759	0.352	-4.997	<0.001
	Morel Var.	-1.711	0.331	-5.166	<0.001
	Clogg Var.	-4.025	0.289	-13.935	<0.001
Age	Standard WEE	0.030	0.026	1.135	0.261
	Independent Var.	0.028	0.022	1.295	0.200
	Morel Var.	0.024	0.023	1.034	0.305
<i>Bias-Reduced</i>	Clogg Var.	0.029	0.027	1.090	0.280
	Standard WEE	-1.552	1.100	-1.410	0.163
	Independent Var.	-1.179	0.855	-1.378	0.173
Female	Morel Var.	-1.154	0.698	-1.653	0.103
	Clogg Var.	-1.489	1.066	-1.400	0.167
	Standard WEE	-0.578	0.854	-0.677	0.500
Comorbid	Independent Var.	-0.727	0.768	-0.946	0.347
	Morel Var.	-0.741	0.707	-1.048	0.298
	Clogg Var.	-0.572	0.877	-0.652	0.517

Table 3

Average relative bias and mean square error of $\hat{\beta}_2$ and empirical coverage probabilities of confidence intervals (percent convergence for WEE) for each simulation specification where the true parameter values are $\beta_2 = \ln(2) \approx 0.69$.

Configuration	Method	Average Relative Bias	Mean Square Error	Empirical Coverage Prob.
$m_i = 5$	WEE	0.659	1.016	0.936
	Independence	0.733	0.323	0.973
	Morel	0.646	0.323	0.968
	Clogg	1.755	0.441	0.973
	WEE	3.474	0.160	0.934
	Independence	1.470	0.139	0.947
	Morel	1.467	0.139	0.944
	Clogg	2.403	0.156	0.950
	WEE	-4.839	0.736	0.928
	Independence	-1.111	0.319	0.970
	Morel	-1.151	0.319	0.965
	Clogg	-3.573	0.427	0.969
$m_i = 10$	WEE	1.431	0.148	0.948
	Independence	1.170	0.127	0.964
	Morel	1.171	0.127	0.960
	Clogg	0.491	0.145	0.962
	WEE	2.656	0.184	0.931
	Independence	-0.375	0.159	0.968
	Morel	-0.389	0.159	0.968
	Clogg	1.494	0.179	0.964
	WEE	1.646	0.066	0.948
	Independence	0.129	0.061	0.966
	Morel	0.128	0.061	0.966
	Clogg	1.098	0.065	0.968
$\phi = 0.9$	WEE	-2.123	0.177	0.940
	Independence	-2.242	0.153	0.980
	Morel	-2.237	0.153	0.981

Configuration	Method	Average Relative Bias	Mean Square Error	Empirical Coverage Prob.
n = 80	Clogg	-3.072	0.174	0.980
	WEE	-0.441	0.074	0.942
	Independence	-0.583	0.069	0.959
	Morel	-0.585	0.069	0.958
	Clogg	-0.919	0.073	0.960

Based on 2000 replications for each simulation for varying levels of the number of observations in each cluster m , levels of ϕ for the bridge distribution, and number of clusters sampled, n .

Table 4

Average relative bias and mean square error of $\hat{\beta}_2$ and empirical coverage probabilities of confidence intervals (percent convergence for WEE) for each simulation specification where the true parameter values are $\beta_2 = \ln(4) \approx 1.39$.

Configuration	Method	Average Relative Bias	Mean Square Error	Empirical Coverage Prob.
$m_i = 5$	WEE	13.229	0.616	0.919
	Independence	2.638	0.413	0.975
	Morel	2.550	0.412	0.970
	Clogg	9.436	0.549	0.966
	WEE	3.802	0.191	0.931
	Independence	-0.623	0.162	0.957
	Morel	-0.624	0.162	0.954
	Clogg	2.381	0.183	0.958
	WEE	8.765	0.845	0.922
	Independence	0.145	0.367	0.970
	Morel	0.064	0.366	0.966
	Clogg	6.331	0.493	0.968
$m_i = 10$	WEE	4.074	0.164	0.946
	Independence	-0.415	0.139	0.962
	Morel	-0.423	0.139	0.958
	Clogg	2.624	0.157	0.964
	WEE	4.277	0.199	0.934
	Independence	0.152	0.169	0.968
	Morel	0.147	0.170	0.971
	Clogg	2.905	0.191	0.968
	WEE	-0.409	0.079	0.943
	Independence	-2.317	0.075	0.956
	Morel	-2.316	0.075	0.956
	Clogg	-1.019	0.078	0.962
$\phi = 0.9$	WEE	3.858	0.184	0.934
	Independence	-0.084	0.157	0.972
	Morel	-0.095	0.157	0.972

Configuration	Method	Average Relative Bias	Mean Square Error	Empirical Coverage Prob.
n = 80	Clogg	2.577	0.178	0.971
	WEE	1.243	0.071	0.948
	Independence	−0.642	0.066	0.964
	Morel	−0.646	0.066	0.964
	Clogg	0.661	0.070	0.964

Based on 2000 replications for each simulation for varying levels of the number of observations in each cluster m , levels of ϕ for the bridge distribution, and number of clusters sampled, n .

Table 5

Average relative bias and mean square error of $\hat{\beta}_2$ and empirical coverage probabilities of confidence intervals (percent convergence for WEE) for each simulation specification where the true parameter values are $\beta_2 = \ln(16) \approx 2.77$.

Configuration	Method	Average Relative Bias	Mean Square Error	Empirical Coverage Prob.
$m_i = 5$	WEE (71.5%)	201.601	284.104	0.709
	Independence	-0.898	1.710	0.942
	Morel	-2.725	1.536	0.925
	Clogg	13.976	2.450	0.960
	WEE (97.0%)	26.499	10.993	0.902
	Independence	-1.324	0.656	0.936
	Morel	-1.567	0.651	0.929
	Clogg	4.303	0.732	0.932
	WEE (85.2%)	101.667	95.892	0.821
	Independence	2.510	1.644	0.960
	Morel	1.210	1.515	0.942
	Clogg	17.300	2.200	0.952
$\phi = 0.9$	WEE (98.6%)	16.440	4.266	0.912
	Independence	-1.021	0.442	0.940
	Morel	-1.149	0.440	0.926
	Clogg	5.044	0.530	0.944
$m_i = 10$	WEE (97.3%)	22.276	8.897	0.912
	Independence	-2.233	0.638	0.952
	Morel	-2.539	0.631	0.943
	Clogg	3.295	0.718	0.964
	WEE	4.517	0.370	0.941
	Independence	-1.805	0.273	0.940
	Morel	-1.934	0.272	0.932
	Clogg	0.618	0.282	0.957
	WEE (99.8%)	10.399	0.700	0.925
	Independence	-0.878	0.418	0.966
	Morel	-0.952	0.418	0.962

Configuration	Method	Average Relative Bias	Mean Square Error	Empirical Coverage Prob.
n = 80	Clogg	5.255	0.518	0.969
	WEE	3.126	0.207	0.936
	Independence	-1.669	0.173	0.947
	Morel	-1.685	0.173	0.944
	Clogg	1.062	0.185	0.956

Based on 2000 replications for each simulation for varying levels of the number of observations in each cluster m , levels of ϕ for the bridge distribution, and number of clusters sampled, n .