Published in final edited form as:

Stat Med. 2015 February 10; 34(3): 444–453. doi:10.1002/sim.6348.

Using the jackknife for estimation in log link Bernoulli regression models

Stuart R. Lipsitz^{a,*}, Garrett M. Fitzmaurice^b, Alex Arriaga^a, Debajyoti Sinha^c, and Atul A. Gawande^a

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

^bHarvard Medical School, Boston, MA, U.S.A

^cFlorida State University, Tallahassee, FL, U.S.A

Abstract

Bernoulli (or binomial) regression using a generalized linear model with a log link function, where the exponentiated regression parameters have interpretation as relative risks, is often more appropriate than logistic regression for prospective studies with common outcomes. In particular, many researchers regard relative risks to be more intuitively interpretable than odds ratios. However, for the log link, when the outcome is very prevalent, the likelihood may not have a unique maximum. To circumvent this problem, a 'COPY method' has been proposed, which is equivalent to creating for each subject an additional observation with the same covariates except the response variable has the outcome values interchanged (1's changed to 0's and 0's changed to 1's). The original response is given weight close to 1, while the new observation is given a positive weight close to 0; this approach always leads to convergence of the maximum likelihood algorithm, except for problems with convergence due to multicollinearity among covariates. Even though this method produces a unique maximum, when the outcome is very prevalent, and/or the sample size is relatively small, the COPY method can yield biased estimates. Here, we propose using the jackknife as a bias-reduction approach for the COPY method. The proposed method is motivated by a study of patients undergoing colorectal cancer surgery.

Keywords

bias-reduction; copy method; maximum likelihood; non-convergence; relative risk regression

1. Introduction

Bernoulli (dichotomous/binary) outcomes commonly arise in studies in the health, behavioral, and social sciences. Logistic regression is arguably the most widely used model for relating a Bernoulli outcome to a set of covariates. However, because the odds ratios (ORs) in a logistic regression are not directly interpretable as relative risks (RRs) unless the outcome is rare, a Bernoulli regression with a log link function [1] is a more appealing

*Correspondence to: Stuart R. Lipsitz, Brigham and Women's Hospital, Boston, MA, U.S.A.

†slipsitz@partners.org

choice for many cross-sectional and prospective studies. For example, in clinical trials, the RR is the preferred metric for reporting treatment differences in a binary outcome; in many cases, the reported RR must be adjusted for variables used in the study design (e.g., blocking factors) and/or baseline variables that are highly unbalanced between treatment groups. Also, in many public health applications, risk and prevalence are the quantities of main interest, and the RR (or prevalence ratio) provides essential information about the association with an exposure of interest that can be used for planning health policy/services or for targeting subpopulations for health interventions. When binary outcomes are analyzed using the generalized linear model with a log link function, one can easily show that the regression parameters have interpretation as log-RRs.

In general, regardless of the choice of link function relating the Bernoulli outcome to the covariates, maximum likelihood (ML) tends to be the method of choice for estimating the Bernoulli regression model parameters. However, because the probabilities are not constrained to be less than 1 in log link regression, when the binary outcome is very prevalent the likelihood may not have a unique maximum, even in large samples. To circumvent this issue of a non-unique maximum for log link regression with Bernoulli outcomes, Deddens *et al.* [2] proposed the ‘COPY method’. The COPY method creates an additional observation for each subject. Specifically, with n independent subjects, if we let Y_i be the Bernoulli (0/1) outcome for subject i , the COPY method creates an additional response $(1 - Y_i)$ for subject i with the same covariates [3]. The original observation (Y_i) is given weight 1, and the new observation $(1 - Y_i)$ is given a small positive weight of order n^{-1} ; both sets of observations are included in the log link regression model. Treating these new observations as independent of the original data and forming a weighted Bernoulli likelihood, Savu *et al.* [4] showed that a unique maximum exists for the weighted likelihood because the weights are positive for both the original Y_i and the additional $(1 - Y_i)$ (this assumes that the usual issues due to multicollinearity that could cause non-uniqueness are not present).

Although the COPY method circumvents the uniqueness problem of the maximum likelihood estimator (MLE), the finite sample bias of the approach has not been studied. In particular, when the binary outcome is very prevalent, and/or the sample size is relatively small, the COPY method may yield biased estimates of the regression parameters. Here, we propose the use of the jackknife [5] to reduce the finite sample bias of the COPY estimator. Specifically, the jackknife has been shown [6] to remove the leading term in the asymptotic bias of the solution to an asymptotically unbiased estimating equation.

Our paper is motivated by a study of best practices for patients undergoing surgery for colorectal cancer [7]. A panel of colorectal and general surgeons was assembled to ascertain a set of 37 evidence-based practices that they felt were the most pertinent to the evaluation and management of a patient undergoing colorectal surgery after the diagnosis of colon cancer. For patients with low-lying rectal cancer in particular, low anterior resection (LAR) surgery to remove the cancer is the treatment of choice. Following the LAR surgery, the patient receives a stoma to collect fecal matter in an external pouch (to be disposed of externally). One of the key best practices for patients receiving LAR surgery is patient consultation with an ostomy nurse prior to surgery. The ostomy nurse cares for patients with

ostomies to ensure the correct cleansing is carried out in order to avoid infection and other complications post-surgery. The medical records of 102 low-lying rectal cancer patients from three hospitals were reviewed (due to confidentiality, hospital names must remain anonymous). Here, the binary outcome of interest is whether an ostomy nurse was consulted (yes, no) prior to surgery. An ostomy nurse was not consulted in 48 out of the 102 (47.1%) patients, indicating that doctors are not meeting the guideline for approximately half of the patients. In this study, the main predictors of interest are the patient's creatinine level (in mmol/L), age (in years), and American Society of Anesthesiologists' (ASA) score; the latter is a global assessment of the physical status of the patient [8]. The ASA score yielded a two-level indicator of a patient's preoperative disease status at diagnosis (0 = mild disease and 1 = life-threatening disease). Higher creatinine level has been shown to be associated with increased postoperative morbidity and mortality [9,10]. Table I presents the overall distributions of ostomy nurse consultations, ASA score, age, and creatinine levels for the 102 patients. Although all patients should have ostomy nurse consultations, it is of interest to explore whether preoperative disease status (ASA score), age, and creatinine level are predictive of patients who will have ostomy nurse consultations. We would expect that patients with higher risks of complications (i.e., patients who are older, with worse ASA score, and higher creatinine levels) to have ostomy nurse consultations.

In Section 2, we briefly describe the underlying Bernoulli likelihood for regression models with the log link. In Section 3, we discuss the COPY approach [2,4]; in Section 4, we describe the jackknife bias correction applied to the COPY estimator. In Section 5, we apply the bias correction approaches to the analyses of predictors of whether or not an ostomy nurse is consulted for patients undergoing surgery for newly diagnosed colorectal cancer [7]. Finally, in Section 6, we present results of a simulation study of bias correction for the log link. One additional approach proposed to obtain consistent estimates of the log-linear regression parameters is a quasi-likelihood, method of moments, technique in which the Bernoulli outcome is naively assumed to have a Poisson distribution [11–13]; we compare the Poisson regression estimator to the others in both the analyses of the colorectal cancer dataset and in the simulations.

2. Bernoulli likelihood

Let Y_i denote the binary response for subject i , $i = 1, \dots, n$, where n is the number of independent subjects. Let $E[Y_i | \mathbf{x}_i, \boldsymbol{\beta}] = P(Y_i = 1 | \mathbf{x}_i, \boldsymbol{\beta}) = p_i$ be the success probability, where $\mathbf{x}_i$ is a $(K \times 1)$ vector of covariates and $\boldsymbol{\beta}$ is a vector of regression coefficients. The success probability is related to the linear predictor $\eta_i = \mathbf{x}_i' \boldsymbol{\beta}$ through the log link function $\log(p_i) = \eta_i = \mathbf{x}_i' \boldsymbol{\beta}$, or equivalently,

$$p_i = \exp(\eta_i) = \exp(\mathbf{x}_i' \boldsymbol{\beta}).$$

One can easily show that the elements of $\boldsymbol{\beta}$ have interpretation as log-RRs. For the remainder of this paper, we assume that the main interest centers on estimating the regression parameter vector $\boldsymbol{\beta}$.

The Bernoulli likelihood is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n p_i^{Y_i} (1 - p_i)^{1-Y_i}. \quad (1)$$

In general, to estimate $\boldsymbol{\beta}$, the ML estimating equations are $U(\hat{\boldsymbol{\beta}}) = 0$, where

$$U(\boldsymbol{\beta}) = \sum_{i=1}^n \frac{d_i \mathbf{X}_i (Y_i - p_i)}{p_i(1 - p_i)}, \quad (2)$$

with $d_i = \frac{\partial p_i}{\partial \eta_i}$. For the log link, $d_i = p_i$ so that

$$U(\boldsymbol{\beta}) = \frac{\partial \log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{x_i (Y_i - p_i)}{1 - p_i}. \quad (3)$$

The MLE is the asymptotically efficient estimate.

It has been noted by several authors that maximization of the Bernoulli likelihood for models with a log link is problematic; specifically, convergence problems in the Newton-Raphson algorithm are common when the success probabilities are close to one [1, 12, 13], because the denominator $(1 - p_i)$ of (3) approaches 0, and the log link does not have natural constraints forcing $(1 - p_i)$ to be greater than 0. That is, a specific feature of the log link model is that the model respects the natural constraints on probabilities, ensuring that predicted probabilities are mapped to the $[0, 1]$ range, only if the regression parameters satisfy the constraints:

$$\mathbf{x}_i' \boldsymbol{\beta} \leq 0, \text{ for all } 1 \leq i \leq n.$$

The MLE is obtained by maximization of the likelihood over this restricted parameter space. In practice, convergence problems with the Newton-Raphson algorithms implemented in standard commercial software for generalized linear models, for example, SAS, Stata, and SPSS, are common when fitting the log link binomial model in datasets arising from large values of p_i [13].

3. COPY estimator

The COPY method [2, 4] requires creation of an additional response with the outcome status reversed, $(1 - Y_i)$, for each subject with the same covariates, and assigns weights to the Y_i and $(1 - Y_i)$. In general, the original Y_i is assigned weight $(c - 1)/c$, and $(1 - Y_i)$ is assigned weight $1/c$, where c is the number of ‘copies’. Without loss of generality, by multiplying these weights by the constant $c/(c - 1)$, we obtain equivalent estimates by assigning Y_i weight 1 and $(1 - Y_i)$ weight $1/(c - 1)$.

Using both sets of observations, the COPY method then maximizes an approximate log likelihood that is a weighted sum of the usual data log likelihood and the log likelihood of the data with outcome status reversed, with the relative weight of the former being very large (i.e., $1/(c-1) \ll 1$). As a result, the COPY estimator inherits all of the properties of the ML estimator of the log link regression, including consistency and asymptotic normality properties (under regularity assumptions [14]). Specifically, Savu *et al.* [4] show that if c is selected to converge to infinity at an appropriate rate with the sample size n , then the COPY estimator converges to the true parameters of the log link regression model and has an asymptotic normal distribution.

To use the jackknife to reduce the finite sample bias (discussed in greater detail in the next section), $(c-1)$ must be at least of order n , suggesting a choice of $c \geq n+1$. We propose the use of $c = n+1$, so that Y_i is assigned weight 1 and $(1-Y_i)$ is assigned weight $1/n$. With these weights, the bias of the COPY estimator will be of order $1/n$, so that the bias-corrected jackknife COPY estimator, which reduces the first-order bias, should have minimal bias. In the COPY method, the sum of the weights equals $n+1$, so this approach effectively adds 1 observation to the full dataset. In contrast, Deddens *et al.* [2] used $c = 1000$ in their examples; however, the jackknife should reduce the first-order bias as long as $c \geq n+1$ as this ensures the COPY estimator inherits the properties of the ML estimator and that the leading term in the bias is of order $1/n$.

Using the results of Savu *et al.* [4], provided that Y_i and $(1-Y_i)$ have positive weights, the weighted ML estimator of β will be unique for the log link regression. In general, n^{-1} times the score function for the COPY method can be written as

$$n^{-1}U(\beta) = n^{-1} \sum_{i=1}^n \frac{x_i(Y_i - p_i)}{1 - p_i} + n^{-2} \sum_{i=1}^n \frac{x_i(1 - Y_i - p_i)}{1 - p_i}. \quad (4)$$

The second term in (4) converges to 0 at $o(1)$, and the COPY score function is asymptotically equivalent to (2), so that the COPY estimator is consistent, asymptotically unbiased, and asymptotically equivalent to the MLE.

4. COPY estimator with jackknife bias correction

The jackknife procedure that we propose using is derived by deleting each subject, one at a time, and recomputing the estimate of β . Let $\hat{\beta}_{-i}$ be the COPY estimate obtained after deleting the i th individual. We use the jackknife procedure [5] to estimate the regression parameters and their variances. The jackknife procedure involves the pseudovectors

$$\tilde{\beta}_i = n\hat{\beta} - (n-1)\hat{\beta}_{-i},$$

for $i = 1, \dots, n$. Then, a jackknife estimator of β is

$$\tilde{\beta} = n^{-1} \sum_{i=1}^n \tilde{\beta}_i = n^{-1} \sum_{i=1}^n (n\hat{\beta} - (n-1)\hat{\beta}_{-i}) = n\hat{\beta} - (n-1)\bar{\beta}, \quad (5)$$

where

$$\bar{\beta} = \sum_{i=1}^n \hat{\beta}_{-i}.$$

Note, when calculating $\hat{\beta}_{-i}$ for the COPY method, Y_i is assigned weight 1 and $(1 - Y_i)$ is assigned weight $1/(n-1)$. Because the weights for Y_i and $(1 - Y_i)$ are again positive, using the results of Savu *et al.* [4], the ML estimator $\hat{\beta}_{-i}$ should be unique in any jackknife subsample. Specifically, Savu *et al.* [4] provide a rigorous, theoretical proof that the modified data likelihood has a unique global maximum point and that the global maximum point is situated in the interior of the restricted parameter space (see Appendix B in [4]).

The COPY estimator can be considered an asymptotically unbiased estimating equation for β , and Chen and Hall [6] showed that the jackknife estimator in (5) removes the leading term in the asymptotic bias of the solution to an asymptotically unbiased estimating equation. Thus, our proposed jackknifed COPY estimator should reduce the first order bias. Finally, the jackknife estimator of the variance of $\tilde{\beta}$ is

$$V_J = [(n-1)n]^{-1} \sum_{i=1}^n (\tilde{\beta}_i - \tilde{\beta})(\tilde{\beta}_i - \tilde{\beta})'.$$

5. Application to study of consultation with an ostomy nurse

In this section, we apply the proposed methodology to the analysis of colorectal cancer patients receiving a LAR surgery to remove the cancer. LAR is the treatment of choice. Following the LAR surgery, the patient receives a stoma to collect fecal matter in an external pouch, and one of the key practices for patients receiving LAR surgery is patient consultation (yes/no) with an ostomy nurse prior to surgery. The study includes $n = 102$ colon cancer patients from three hospitals (due to confidentiality, hospital names must remain anonymous) whose medical records were reviewed. The outcome is the Bernoulli variable denoting ‘patient consultation with an ostomy nurse’ (0 = yes, 1 = no). Note that the study investigators were interested in reporting factors associated with the probability of *no* consultation [7]; as a result, we use that convention here to predict the probability of no consultation as a function of patient characteristics. There are three main patient predictors of interest: the two-level ASA disease status at diagnosis (0 = mild disease and 1 = life-threatening disease), patient’s creatinine level (in mmol/L), and age (in years). A priori, our medical colleagues conjectured that if there was any effect at all, that younger patients, patients with better baseline disease (ASA=0), and with lower creatinine levels would be less likely to have a consultation with an ostomy nurse.

To examine the relationship between the binary outcome and these three covariates, we fit the log link regression model,

$$\log(p_i) = \beta_0 + \beta_1 \text{ASA}_i + \beta_2 \text{Age}_i + \beta_3 \text{Creatinine}_i \quad (6)$$

where $\text{ASA}_i = 1$ if life-threatening disease and 0 otherwise; Age_i is age in years (although the effect is multiplied by 100 for easier comparison of results in Table II), Creatinine_i is creatinine level in mmol/L, and p_i is the probability that an ‘ostomy nurse is not consulted’, given the values of these three covariates.

Table II gives the estimates of β obtained using (i) standard ML, as implemented in PROC GENMOD in SAS, (ii) the Poisson quasi-likelihood approach [1,12,13], (iii) the standard COPY approach (without application of the jackknife) with $c = n$ and $c = 1000$ (as suggested by Deddens *et al.* [2]), and (iv) the jackknife COPY estimator with $c = n$ and $c = 1000$. For all methods except standard ML, we wrote macros in SAS (available upon request from the authors). Of note, there were convergence problems with ML but not with any of the other methods. The ML estimates in Table II are from the 100th iteration of the Newton-Raphson algorithm implemented in PROC GENMOD in SAS and are quite different from the estimates yielded by the other approaches; in particular, estimates of the effects of ASA score and age are approximately half the magnitude of the estimates produced by the other methods. In Section 7, we briefly discuss how the convergence problems that arise when fitting the log link binomial model are primarily due to the algorithms or computation methods commonly used for finding the MLE; the results of our simulations (in Section 6) highlight the magnitude of this problem for the Newton-Raphson algorithm implemented in PROC GENMOD in SAS.

There were discernible differences in the estimated regression coefficients for the alternative approaches to ML, especially for the effect of creatinine levels. The effect of creatinine was approximately 40% smaller for the jackknife COPY method ($\tilde{\beta}_3 \approx 0.12$) versus the COPY method ($\hat{\beta}_3 \approx 0.22$); furthermore, the jackknife estimates of the effect of creatinine levels (with $c = n$ and $c = 1000$) are non-significant, whereas the COPY estimates are significant. Unlike the jackknife estimates, the COPY and Poisson methods yield results that are statistically significant but in the opposite direction of what was expected on subject matter grounds—that patients with higher creatinine levels are less likely to have an ostomy nurse consulted. We note that the computational time to run our SAS macro for the jackknife estimator was 1 s on a PC with 32-bit Intel Core2 Duo E8400, 3.0 GHz processor with 4 GB of RAM. Thus, the jackknife COPY estimator is computationally feasible, especially when needed for datasets with a relatively small number of observations.

On the basis of this single example, it is not possible to discern whether the differences observed in Table II are due to biases of one or more of the methods or simply due to sampling variability. Thus, to explore this further, we conducted a simulation study comparing the methods; the results of the simulation study are reported in the next section.

6. Simulation study

Here, we study the finite sample bias in estimating β using standard ML, the Poisson quasi-likelihood approach [1, 12, 13], the standard COPY method (without application of the jackknife) with $c = n$ and $c = 1000$, and the jackknife COPY estimator with $c = n$ and $c = 1000$.

For the first set of simulations, we consider a log regression model with one continuous covariate,

$$\log(p_i) = \beta_0 + \beta_1 x_i,$$

where, for simplicity, we let x_i have a triangular distribution on the interval (0,1) with mode 0.8. We arbitrarily let β_1 be negative, so that $\exp(\beta_0)$ is the maximum value for any p_i and $\exp(\beta_0 + \beta_1)$ is the minimum value for any p_i . Then, to specify the log regression model for the simulations, we first specified the maximum (p_{max}) and minimum (p_{min}) values that p_i can attain, which in turn fully specifies the parameters $\beta_0 = \log(p_{max})$ and $\beta_1 = \log(p_{min}) - \log(p_{max})$. Specification of the simulation model in this way allows us to explore properties of the methods as p_i approaches 1. We note that the choice of sign for β_1 is completely arbitrary although its value determines the minimum and maximum value for any p_i .

For the different values of p_{max} and p_{min} (or equivalently, β_0 and β_1), we conducted simulations for $n = 40$, with 2500 simulation replications performed for each configuration. We first performed simulations with $p_{min} = 0.01$ and $p_{max} = 0.5, 0.6, 0.7$, giving three sets of simulations. Finally, to examine a more extreme configuration, we performed one set of simulations with $p_{min} = 0.6$ and $p_{max} = 0.99$ and one set with $p_{min} = 0.7$ and $p_{max} = 0.99$.

Table III presents the relative biases, the root mean square error, and the coverage probabilities of 95% Wald confidence intervals, as well as the percent converged for β_1 . We see that the percent convergence for ML (as implemented in PROC GENMOD in SAS) is less than 30% for all configurations. The COPY and jackknife approaches converged for all configurations.

Because the ML approach converged less than 30% for all configurations, it is difficult to draw any general conclusions about ML from the results of these simulations. In contrast, the COPY method converged for all simulations, but had relative biases between 10% and 30%. In general, the COPY method with $c = 1000$ performed slightly better than that with $c = n$; however, it still has bias greater than 10%. Even the Poisson quasi-likelihood approach has relative bias greater than 10% for all simulation configurations. In general, the jackknife COPY estimator appears to have the best properties with respect to bias, with all relative biases less than 5%. Thus, in terms of bias, the jackknife COPY estimator appears to be the method of choice in log link regression when the sample size is relatively small. In terms of root mean square error in small samples, the results in Table III suggest that the improvement in bias reduction of the jackknife COPY estimator over the standard COPY estimator comes at the cost of some increase in variability. In addition, the bias correction is

achieved at the cost of some apparent reduction in coverage probability of the Wald confidence intervals.

We conducted a second set of simulations modelled closely on the colorectal cancer surgery data example in Section 5. Specifically, we considered a log regression model with one binary (x_{1i}) and one continuous (x_{2i}) covariate,

$$\log(p_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i},$$

where we let $x_{1i} \sim \text{Bernoulli}(0.5)$ and $x_{2i} \sim \text{Triangular}$, a triangular distribution on the interval $(0,1)$ with mode 0.3; the distribution for x_{1i} is very similar to the distribution of ASA, while the distribution of x_{2i} mimics the positive skewness of creatinine in the colorectal cancer surgery data. We fixed $\beta_1 = -0.5$ and $\beta_2 = 0.2$, similar to the estimated effects of ASA and creatinine in Table II. We simulated binary responses where the overall $P(Y_i = 1) \approx 0.45$, similar to the overall probability that an ‘ostomy nurse is not consulted’ in the example in Section 5. We conducted simulations for $n = 40, 60$, and 80 with 2500 simulation replications performed for each configuration.

Table IV presents the relative biases, the root mean square error, and the coverage probabilities of 95% Wald confidence intervals, as well as the percent converged for β_1 and β_2 . We see that the percent convergence for ML (as implemented in `PROC GENMOD` in SAS) is approximately in the range 60–75% for all configurations. Because the ML approach converged less than 76% for all configurations, the results for this implementation of ML must be cautiously interpreted. With that caveat, the results in Table IV nonetheless suggest that the relative bias of ML for β_2 is in the 55–70% range. In contrast, the Poisson quasi-likelihood and COPY approaches converged for all configurations. The Poisson estimator yielded relative bias less than 6% for β_1 , but had relative bias of 3–13% for β_2 . Similarly, the COPY estimator had bias less than 7% for β_1 , but had relative bias of 4–17% for β_2 ; the bias for β_2 seems to be smaller for $c = 1000$ than for $c = n$. Finally, the jackknife COPY estimator appears to have the best properties with respect to bias, with relative biases less than 6% for β_1 and less than 3% for β_2 . In addition, the coverage probabilities of the jackknife COPY estimator are closer to the nominal level when $c = 1000$.

7. Conclusion

Relative risks are often considered the preferred measures of association in prospective studies, especially when the binary outcome of interest is common. In particular, many researchers regard RRs to be more intuitively interpretable than ORs. Interestingly, in many studies where RRs are the parameters of primary scientific interest, ORs are reported instead. This can be explained in part by the technical advantages of logistic regression (e.g., no constraints on the regression parameters) and the widespread availability of appropriate software. RR regression is a special case of generalized linear models, specifically with a log link function for the binomial (or Bernoulli) outcome. However, for the log link, when the outcome is very prevalent, the likelihood may not have a unique maximum. To circumvent this problem, Deddens *et al.* [2] proposed the ‘COPY method’. Although the method produces a unique maximum, the COPY method can yield biased estimates when the sample

size is relatively small. In this paper, we have described a jackknife approach to bias correction for the COPY method. The approach is based on solid statistical theory for bias reduction [5,6], and has demonstrated good properties in our simulations. Moreover, the jackknife approach is straightforward to implement and is not computationally time consuming for the relatively small datasets that it would be applied to.

The results of the simulations demonstrate that the proposed method reduces the finite sample bias of ML. Because of numerical difficulties associated with their restricted parameter space, RR regressions often fail to converge when ML estimates of the RR parameters are located close to the boundary of the restricted parameter space. As the theory suggests [4], we found the jackknife approach applied to the COPY method always converges and has relatively small bias. Because of the broad range of possible data configurations, it is difficult to draw definitive conclusions from the results of the simulation studies. Nonetheless, on the basis of the simulations reported here, the jackknife COPY method appears to have discernibly less bias than the standard COPY estimator.

Finally, the results of the simulations confirm that convergence problems with the Newton-Raphson algorithms implemented in standard commercial software for generalized linear models, for example, SAS, Stata, and SPSS, are common when fitting the log link binomial model in datasets where the binary outcome is prevalent. Specifically, we found that the percent convergence for ML (as implemented in PROC GENMOD in SAS) was less than 30% in one set of simulation configurations and less than 76% in the second set. However, we must emphasize that the convergence problems that arise when fitting the log link binomial model are largely due to the algorithms or computation methods commonly used for finding the MLE rather than the ML method of estimation *per se*. That is, convergence problems are common when fitting the log link binomial model using these algorithms despite the fact that the likelihood function has a unique finite maximum. More reliable computation methods for finding the MLE for log link binomial models have recently been developed [15] but are not yet widely implemented in standard commercial software for generalized linear models. A topic of future research is to examine the bias reduction properties of the jackknife when applied to ML estimation of the log binomial regression parameters using these more reliable computation methods for finding the MLE.

Acknowledgments

We thank Edward Giovannucci and Caprice Greenberg for the advice on the analysis of the data from the colorectal cancer surgery study. We are grateful for the support provided by grant CA 160679 from the US National Institutes of Health.

References

1. Wacholder S. Binomial regression in GLIM: estimating risk ratios and risk differences. *American Journal of Epidemiology*. 1986; 123:174–184. [PubMed: 3509965]
2. Deddens, JA., Petersen, MR., Lei, X. Estimation of prevalence ratios when PROC GENMOD does not converge. *Proceedings of the 28th Annual SAS Users Group International Conference*; 30 March–2 April 2003; Cary NC: SAS Institute Inc; p. 270-28.(Available from: <http://www2.sas.com/proceedings/sugi28/270-28.pdf>) [Accessed on 11 August 2014]

3. Lumley, T., Kornmal, R., Ma, S. Relative risk regression in medical research: models, contrasts, estimators, and algorithms; UW Biostatistics Working Paper Series. 2006. p. 293(Available from: <http://www.bepress.com/uwbiostat/paper293>) [Accessed on 11 August 2014]
4. Savu A, Liu Q, Yasui Y. Estimation of relative risk and prevalence ratio. *Statistics In Medicine*. 2010; 29:2269–2281. [PubMed: 20564738]
5. Quenouille MH. Notes on bias in estimation. *Biometrika*. 1956; 61:353–360.
6. Chen SX, Hall P. Effects of bagging and bias correction on estimators defined by estimating equations. *Statistica Sinica*. 2003; 13:97–109.
7. Arriaga AF, Lancaster RT, Berry WR, Regenbogen SE, Lipsitz SR, Kaafarani HM, Elbardissi AW, Desai P, Ferzoco SJ, Bleday R, Breen E, Kastrinakis WV, Rubin MS, Gawande AA. The better colectomy project: association of evidence-based best-practice adherence rates to outcomes in colorectal surgery. *Annals of Surgery*. 2011; 250:507–513.
8. Owens WD, Felts JA, Spitznagel EL Jr. ASA physical status classifications: a study of consistency of ratings. *Anesthesiology*. 1978; 49:239–243. [PubMed: 697077]
9. Kozol RA, Hyman N, Strong S, Whelan RL, Cha C, Longo WE. Minimizing risk in colon and rectal surgery. *American Journal of Surgery*. 2007; 194:576–587. [PubMed: 17936417]
10. Kirchhoff P, Clavien PA, Hahnloser D. Complications in colorectal surgery: risk factors and preventive strategies. *Patient Safety in Surgery*. 2010; 4:5. [PubMed: 20338045]
11. McNutt LA, Wu C, Xue X, Hafner JP. Estimating relative risk in cohort studies and clinical trials of common events. *American Journal of Epidemiology*. 2003; 157:940–943. [PubMed: 12746247]
12. Zou G. A modified Poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology*. 2004; 159:702–706. [PubMed: 15033648]
13. Carter RE, Lipsitz SR, Tilley BC. Quasi-likelihood estimation for relative risk regression models. *Biostatistics*. 2005; 6:39–44. [PubMed: 15618526]
14. Fahrmeir L, Kaufmann H. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics*. 1985; 13:342–368.
15. Marschner IC, Gillett AC. Relative risk regression: reliable and flexible methods for log binomial models. *Biostatistics*. 2012; 13:179–199. [PubMed: 21914729]

Table I

Overall patient characteristics.

Variable	Level	Number (%)
Patient consultation	No	48 (47.1%)
with an ostomy nurse	Yes	54 (52.9%)
ASA score	0	49 (48.0%)
	1	53 (52.0%)
Age, median (range)		62 (19–95)
Creatinine (mmol/L), median (range)		0.9 (0.04–3.2)

Comparison of log link regression estimates for the probability that an ostomy nurse is not consulted.

Table II

Effect	Approach	Estimate	SE	Z-statistic	P-value
Intercept	Standard ML	-0.496	0.333	-1.49	0.136
	Poisson	-0.296	0.437	-0.68	0.500
	COPY ($c = 102$)	-0.239	0.597	-0.40	0.688
	COPY ($c = 1000$)	-0.226	0.608	-0.37	0.710
ASA	COPY with jackknife($c = 102$)	-0.192	0.685	-0.28	0.779
	COPY with jackknife($c = 1000$)	-0.219	0.723	-0.30	0.762
	Standard ML	0.238	0.144	1.65	0.101
	Poisson	0.574	0.241	2.38	0.019
Age ($\times 100$)	COPY ($c = 102$)	-0.560	0.253	-2.21	0.027
	COPY ($c = 1000$)	-0.572	0.256	-2.23	0.026
	COPY with jackknife($c = 102$)	-0.555	0.252	-2.20	0.028
	COPY with jackknife($c = 1000$)	-0.546	0.261	-2.09	0.036
Creatinine	Standard ML	-0.360	0.500	-0.72	0.465
	Poisson	-0.844	0.645	-1.31	0.194
	COPY ($c = 102$)	-0.810	0.830	-0.98	0.332
	COPY ($c = 1000$)	-0.820	0.850	-0.96	0.332
Creatinine	COPY with jackknife($c = 102$)	-0.703	0.977	-0.72	0.472
	COPY with jackknife($c = 1000$)	-0.656	1.031	-0.64	0.525
	Standard ML	0.145	0.230	0.63	0.528
	Poisson	0.301	0.108	2.80	0.006
Creatinine	COPY ($c = 102$)	0.217	0.065	3.33	0.001
	COPY ($c = 1000$)	0.217	0.063	3.46	0.001
	COPY with jackknife($c = 102$)	0.123	0.108	1.14	0.254
	COPY with jackknife($c = 1000$)	0.119	0.112	1.06	0.291

ML, maximum likelihood.

Table III

Simulation results for estimation of β_1 for $\log(p_i) = \beta_0 + \beta_1 x_i$, $x_i \sim \text{Triangular}$, and $n = 40$. Here, $\beta_0 = \log(p_{\max})$ and $\beta_1 = \log(p_{\min}) - \log(p_{\max})$, where, for a given simulation configuration, $p_{\max}$ and $p_{\min}$ are the specified maximum and minimum values of p_i , respectively.

$p_{\min}$	$p_{\max}$	Method	Percent relative bias of β_1	Root MSE	Coverage probability	Percent converged
0.01	0.50	Standard ML [*]	-94.34	47.16	100.0%	0.3 %
		Poisson	46.40	193.22	81.9%	100.0 %
		COPY($c = 40$)	-26.14	23.70	96.2 %	100.0 %
		COPY with jackknife($c = 40$)	3.48	36.59	91.1 %	100.0 %
		COPY($c = 1000$)	16.80	46.75	97.5 %	100.0 %
		COPY with jackknife($c = 1000$)	-4.97	86.84	91.0 %	100.0 %
0.01	0.60	Standard ML	15.63	47.54	100.0%	9.0 %
		Poisson	23.80	113.05	90.8 %	100.0 %
		COPY ($c = 40$)	-28.77	23.80	97.6 %	100.0 %
		COPY with jackknife($c = 40$)	2.59	29.79	91.9 %	100.0 %
		COPY($c = 1000$)	16.87	42.74	98.4 %	100.0 %
		COPY with jackknife($c = 1000$)	1.52	81.42	93.6 %	100.0 %
0.01	0.70	Standard ML	-16.49	37.74	97.8 %	3.6 %
		Poisson	71.09	250.89	92.7 %	100.0 %
		COPY($c = 40$)	-18.92	18.27	93.8 %	100.0 %
		COPY with jackknife($c = 40$)	2.72	22.59	92.5 %	100.0 %
		COPY($c = 1000$)	15.45	59.67	98.8 %	100.0 %
		COPY with jackknife($c = 1000$)	2.43	24.80	93.2 %	100.0 %
0.60	0.99	Standard ML	-80.63	5.23	97.9 %	27.1 %
		Poisson	10.47	4.16	96.9 %	100.0 %
		COPY($c = 40$)	-11.57	4.26	94.1 %	100.0 %
		COPY with jackknife($c = 40$)	2.71	5.32	96.7 %	100.0 %
		COPY($c = 1000$)	-12.37	2.86	95.8 %	100.0 %
		COPY with jackknife($c = 1000$)	1.11	4.69	96.6 %	100.0 %
0.70	0.99	Standard ML	-84.75	3.59	99.1 %	10.8%
		Poisson	10.84	18.65	94.4 %	100.0 %
		COPY($c = 40$)	-19.13	2.22	97.3 %	100.0 %

P_{min}	P_{max}	Method	Percent relative bias of $\hat{\theta}_1$	Root MSE	Coverage probability	Percent converged
		COPY with jackknife($c = 40$)	-5.15	2.44	95.1 %	100.0 %
		COPY($c = 1000$)	-15.78	3.06	97.8 %	100.0 %
		COPY with jackknife($c = 1000$)	-2.35	4.09	95.1 %	100.0 %

[‡]Standard maximum likelihood (ML) as implemented in PROC GENMOD in SAS.

Table IV

Simulation results for estimation of β_1 and β_2 for $\log(p_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$, $x_{1i} \sim \text{Bernoulli}(0.5)$, $x_{2i} \sim \text{Triangular}$, and $n = 40, 60, 80$.

n	β	Method	Percent relative bias of β	Root MSE	Coverage probability	Percent converged
40	β_1	Standard ML ^s	-16.45	4.23	94.5 %	62.9 %
		Poisson	1.30	4.44	96.7 %	100.0 %
		COPY($c=40$)	-5.65	4.10	95.2 %	100.0 %
		COPY with jackknife($c=40$)	-6.21	4.11	96.0 %	100.0 %
		COPY($c=1000$)	7.10	4.55	96.6 %	100.0 %
	β_2	COPY with jackknife($c=1000$)	-4.26	4.23	98.3 %	100.0 %
		Standard ML	-68.34	7.34	96.8 %	62.9 %
		Poisson	-12.77	9.19	96.0 %	100.0 %
		COPY($c=40$)	-17.41	8.67	89.4 %	100.0 %
		COPY with jackknife($c=40$)	-0.95	9.86	91.2 %	100.0 %
60	β_1	COPY($c=1000$)	-15.09	9.07	87.2 %	100.0 %
		COPY with jackknife($c=1000$)	-1.55	9.98	93.4 %	100.0 %
		Standard ML	-6.69	3.25	96.0 %	67.5 %
		Poisson	5.45	3.38	97.1 %	100.0 %
		COPY($c=60$)	1.02	3.22	96.4 %	100.0 %
	β_2	COPY with jackknife($c=60$)	0.89	3.25	96.8 %	100.0 %
		COPY($c=1000$)	4.89	3.54	96.0 %	100.0 %
		COPY with jackknife($c=1000$)	-0.79	3.33	97.8 %	100.0 %
		Standard ML	-60.99	6.04	97.4 %	67.5 %
		Poisson	-5.68	7.59	94.4 %	100.0 %
80	β_1	COPY($c=60$)	-9.23	7.18	89.6 %	100.0 %
		COPY with jackknife($c=60$)	2.52	7.75	91.8 %	100.0 %
		COPY($c=1000$)	-8.22	7.34	88.4 %	100.0 %
		COPY with jackknife($c=1000$)	1.51	7.86	93.2 %	100.0 %
		Standard ML	-6.02	2.79	95.2 %	75.3 %
		Poisson	1.66	2.89	96.2 %	100.0 %
		COPY($c=80$)	-1.65	2.79	95.8 %	100.0 %
		COPY with jackknife($c=80$)	-1.96	2.80	96.0 %	100.0 %

n	β	Method	Percent relative bias of β	Root MSE	Coverage probability	Percent converged
80	β_2	COPY($c = 1000$)	1.31	2.89	95.8 %	100.0%
		COPY with jackknife($c = 1000$)	-2.59	2.78	97.31%	100.0%
		Standard ML	-54.14	5.18	97.2 %	75.3 %
		Poisson	-2.74	6.20	95.6 %	100.0%
		COPY($c = 80$)	-6.55	5.92	91.8%	100.0%
		COPY with jackknife($c = 80$)	-0.27	6.11	93.8 %	100.0%
		COPY($c = 1000$)	-3.98	6.10	91.0 %	100.0%
		COPY with jackknife($c = 1000$)	-0.04	6.18	94.8 %	100.0%

β_2 Standard maximum likelihood (ML) as implemented in PROC GENMOD in SAS.