Biostatistics (2014), **15**, 4, *pp.* 745–756
doi:10.1093/biostatistics/kxu012
Advance Access publication on April 4, 2014

Almost efficient estimation of relative risk regression

GARRETT M. FITZMAURICE*
Harvard Medical School, Boston, MA 02115, USA
fitzmaur@hsph.harvard.edu
STUART R. LIPSITZ, ALEX ARRIAGA
Brigham and Women's Hospital, Boston, MA 02115, USA
DEBAJYOTI SINHA
Florida State University, Tallahassee, FL 32306, USA
CAPRICE GREENBERG
University of Wisconsin, Madison, WI 53792, USA
ATUL A. GAWANDE
Brigham and Women's Hospital, Boston, MA 02115, USA
Ariadne Labs, Boston, MA 02115, USA

Summary

Relative risks (RRs) 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 odds ratios. Although RR regression is a special case of generalized linear models, specifically with a log link function for the binomial (or Bernoulli) outcome, the resulting log-binomial regression does not respect the natural parameter constraints. Because log-binomial regression does not ensure that predicted probabilities are mapped to the [0,1] range, maximum likelihood (ML) estimation is often subject to numerical instability that leads to convergence problems. To circumvent these problems, a number of alternative approaches for estimating RR regression parameters have been proposed. One approach that has been widely studied is the use of Poisson regression estimating equations. The estimating equations for Poisson regression yield consistent, albeit inefficient, estimators of the RR regression parameters. We consider the relative efficiency of the Poisson regression estimator and develop an alternative, almost efficient estimator for the RR regression parameters. The proposed method uses near-optimal weights based on a Maclaurin series (Taylor series expanded around zero) approximation to the true Bernoulli or binomial weight function. This yields an almost efficient estimator while avoiding convergence problems. We examine the asymptotic relative efficiency of the proposed estimator for an increase in the number of terms in the series. Using simulations, we demonstrate the potential for

*To whom correspondence should be addressed.

convergence problems with standard ML estimation of the log-binomial regression model and illustrate how this is overcome using the proposed estimator. We apply the proposed estimator to a study of predictors of pre-operative use of beta blockers among patients undergoing colorectal surgery after diagnosis of colon cancer.

Keywords: Bernoulli likelihood; Convergence problems; Maclaurin series; Poisson regression; Quasi-likelihood.

1. INTRODUCTION

We consider prospective study designs where it is of scientific interest to estimate relative risks (RRs) conditional on covariates. Interestingly, in many studies where RRs are the parameters of primary scientific interest, odds ratios (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. Although RR regression is a special case of generalized linear models, specifically with a log link function for the binomial (or Bernoulli) outcome, the resulting log-binomial regression does not respect the natural parameter constraints. Because log-binomial regression does not ensure that predicted probabilities are mapped to the [0,1] range, maximum likelihood (ML) estimation is often subject to numerical instability that leads to convergence problems. It has been noted by several authors that convergence problems are especially likely to arise when the predicted probabilities are close to 1 (Wacholder, 1986; Lu and Tilley, 2001; Zou, 2004; Carter *and others*, 2005). Several approaches have been proposed to circumvent the convergence problems associated with ML estimation of log-binomial regression. These include: (1) directly estimating RR using a constrained ML method that truncates the range of the probabilities (Wacholder, 1986); (2) adding a small constant to each subject's Bernoulli outcome in the sample (Clogg *and others*, 1991; Deddens *and others*, 2003); (3) indirectly estimating RR using the mathematical relationship between OR and RR for a single binary covariate (Zhang and Yu, 1998); and (4) quasi-likelihood method of moments techniques (Traissac *and others*, 1999; McNutt *and others*, 2003; Zou, 2004; Carter *and others*, 2005).

One approach that has been widely studied is the use of Poisson regression estimating equations (Traissac *and others*, 1999; McNutt *and others*, 2003; Zou, 2004; Carter *and others*, 2005). That is, the Poisson likelihood equations are used to estimate the RR regression parameters without constraints. The estimating equations for Poisson regression yield consistent, asymptotically normal (CAN) estimators of the RR regression parameters. Moreover, Carter *and others* (2005) found that they alleviate convergence problems associated with ML estimation of log-binomial regression parameters. Although estimating equations for Poisson regression yield consistent estimators of the RR regression parameters, they are inefficient because the "weight function" is misspecified as Poisson instead of bionomial or Bernoulli. As we demonstrate later, the loss of efficiency tends to be greatest in the very setting where their use is required, i.e. when predicted probabilities are far from zero.

We consider the relative efficiency of the Poisson regression estimator and develop an alternative, almost efficient estimator for the RR regression parameters. The proposed method uses near-optimal weights based on a Maclaurin series (Taylor series expanded around 0) approximation to the true Bernoulli weight function. If the Maclaurin series is truncated at its first term, this yields the Poisson regression estimator. Truncation at higher terms in the series yields near-optimal weights and almost efficient estimators. Using method-of-moments, assuming the RR regression model is correctly specified, the estimators are consistent for any given truncation of the Maclaurin series approximation to the optimal weight function. We examine the asymptotic relative efficiency (ARE) for an increase in the number of terms in the series. We also make recommendations for choice of the number of terms to avoid similar finite sample convergence problems as with ML estimation of log-binomial regression parameters. We present results of a simulation study that highlight the potential gains in efficiency in finite samples.

Table 1. *Descriptive statistics for patient characteristics from study of pre-operative use of beta blockers among patients undergoing colorectal surgery*

Variable	Level	Number (%)
Pre-operative beta blockers	Yes	130 (61.9%)
	No	80 (38.1%)
ASA score	0	92 (43.8%)
	1	118 (56.2%)
Gender	Male	98 (46.7%)
	Female	112 (53.3%)
Race	White	182 (86.7%)
	Other	28 (33.3%)
Number of comorbidities	0	30 (14.3%)
	1	62 (29.5%)
	2	58 (27.6%)
	3	36 (17.4%)
	4	17 (8.1%)
	5	4 (1.9%)
	6	3 (1.4%)
Age, median (range)		68.5 (26–95)

The proposed method is motivated by a study of best practices for patients undergoing surgery for colorectal cancer (Arriaga *and others*, 2009). A panel of colorectal and general surgeons was assembled to ascertain a set of 37 evidence-based practices that they considered to be the most pertinent to the evaluation and management of a patient undergoing colorectal surgery after diagnosis of colon cancer. Patients with known heart disease who are given beta blockers prior to surgery have been found to have a significantly reduced risk of post-operative death (Poldermans *and others*, 1999). Thus, one of the key practices is giving beta blockers when indications (heart disease) are present. In this study, the medical records of 210 cancer patients with cardiac conditions from three hospitals were reviewed (due to confidentiality, hospital names must remain anonymous). Here the binary outcome of interest is whether the patient was given beta blockers (yes, no) prior to surgery. Upon review of the medical records, it was found that beta blockers were given to 130 out of the 210 (61.9%) patients, indicating that many doctors are not meeting the best practices guideline for almost 40% of the patients. The goal of this study is to determine predictors of pre-operative use of beta blockers in these patients. The main predictors of interest are the patient's age (in years), race (categorized as white versus other), gender, number of comorbidities, and "American Society of Anesthesiologists (ASA) score." The ASA score is a global assessment of the physical status of the patient (Owens *and others*, 1978) and yields a two-level indicator of a patient's pre-operative disease status at diagnosis (0 = mild disease, 1 = life-threatening disease). Table 1 presents the overall distributions of beta blocker use, age, race, gender, number of comorbidities, and ASA score. Although all of these cardiac patients should receive pre-operative beta blockers, it is of interest to explore whether pre-operative disease status (ASA score), age, race, gender, and number of comorbidities are predictive of those patients who were given beta blockers. The study investigators conjectured that patients with higher risks of complications, i.e. patients who are older, with worse ASA score and more comorbidities, are more likely to receive pre-operative beta blockers. It is also of interest to examine whether there are any differences by race and gender.

In Section 2, we describe the RR regression model, the corresponding Bernoulli likelihood, and the proposed estimating equations for the RR regression parameters. In Section 3, we present results of a study

examining the ARE of the proposed estimator for increasing number of terms in the series. In Section 4, we present results of a simulation study that demonstrate the potential gains in efficiency in finite samples.

2. RR REGRESSION MODEL

Let Y_i denote the binary response (success or failure) for subject i, $i = 1, \ldots, n$, where n is the number of independent subjects. Then $E[Y_i|x_i] = P(Y_i = 1|x_i) = p_i$ is the success probability, where x_i is a $(K \times 1)$ vector of covariates. In RR regression, the success probability is modeled using the log link,

$$\log(p_i) = x_i'\beta,$$

or, equivalently, $p_i = e^{x_i'\beta}$, where β is the vector of regression parameters. One can easily show that the elements of β (with the exception of the intercept) have interpretation as log-RRs (see, for example, Jewell, 2003). For the remainder of this paper, we assume that the main interest centers on estimating the regression parameter vector β.

The Bernoulli likelihood is

$$L(\beta) = \prod_{i=1}^{n} p_i^{Y_i}(1 - p_i)^{1-Y_i}. \tag{2.1}$$

The ML estimating equations for β are $S(\hat{\beta}) = 0$, where

$$S(\beta) = \frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^{n} \frac{x_i(Y_i - p_i)}{1 - p_i}. \tag{2.2}$$

The MLE is the asymptotically efficient estimate. However, when the success probability approaches 1, the denominator of (2.2) approaches 0, resulting in convergence problems. Wacholder (1986) constrained the likelihood to prevent $(1 - p_i)$ from approaching 0; however, these modifications are still subject to convergence problems (Baumgarten *and others*, 1989).

In general, to estimate β, one can use estimating equations (or quasi-likelihood equations; Wedderburn, 1974) of the form $U(\hat{\beta}) = 0$, where

$$U(\beta) = \sum_{i=1}^{n} w(p_i)x_i(Y_i - p_i), \tag{2.3}$$

with $w(p_i)$ being a "weight" function of p_i. Assuming that the regression model for p_i has been correctly specified, i.e. $E(Y_i - p_i) = 0$, these estimating equations yield CAN estimators of β for any bounded weight function $w(p_i)$ (see, for example, Rotnitzky, 2009). Specifically, it can be shown that the asymptotic distribution of $\widehat{\beta}$, the estimator for β with a particular choice of $w(p_i)$, satisfies

$$\sqrt{n}(\hat{\beta} - \beta) \to N(0, C_\beta), \tag{2.4}$$

where

$$C_\beta = \lim_{n\to\infty} I_0^{-1} I_1 I_0^{-1},$$

$$I_0 = \frac{1}{n}\sum_{i=1}^{n} w(p_i)p_i x_i x_i',$$

and

$$I_1 = \frac{1}{n}\sum_{i=1}^{n} w^2(p_i)\,\mathrm{Var}(Y_i)x_i x_i'.$$

Note that the asymptotically efficient estimate is the MLE, with weight function $w(p_i) = (1 - p_i)^{-1}$, and asymptotic covariance determined by I_0. Consistent estimators of the asymptotic covariance of the estimated regression parameters can be obtained using the empirical estimator of C_β proposed by Huber (1967), White (1982), and Royall (1986). The empirical variance estimator is obtained by evaluating p_i at $p_i(\hat{\beta})$ and substituting $(Y_i - \hat{p}_i)^2$ for $\mathrm{Var}(Y_i)$; this is widely known as the *sandwich* variance estimator.

The use of Poisson regression for estimating β has received much attention recently (e.g. Traissac *and others*, 1999; McNutt *and others*, 2003; Zou, 2004; Carter *and others*, 2005). The Poisson regression estimating equations are $S_2(\hat{\beta}) = 0$, where

$$S_2(\beta) = \sum_{i=1}^{n} x_i(Y_i - p_i). \tag{2.5}$$

It is apparent that the Poisson regression estimating equations are simply a special case of (2.3) with $w(p_i) = 1$. Thus, although the Poisson regression estimating equations produce consistent estimators of β, they can be quite inefficient because $w(p_i) = 1$ is not the optimal or asymptotically efficient weight function. In general, weight functions closer to $w(p_i) = (1 - p_i)^{-1}$ will have higher efficiency. Thus, the goal of this paper is to choose a $w(p_i)$ close to $(1 - p_i)^{-1}$, but one that also avoids the convergence problems associated with ML. Lumley *and others* (2006) considered the weights $w(p_i) = 1$, $w(p_i) = p_i$, and $w(p_i) = (1 + p_i)^{-1}$ and estimated their relative efficiency (relative to the MLE). Lumley *and others* found that with p_i close to 1 these three estimating equations can give inefficient estimates; this is due to the fact that the three weight functions considered do not closely approximate the optimal weight function.

To develop more efficient estimating equations, we first note that the Maclaurin series (Abramowitz and Stegun, 1970) of $(1 - p_i)^{-1}$ is

$$\frac{1}{1 - p_i} = 1 + p_i + p_i^2 + p_i^3 + p_i^4 + \cdots = \sum_{m=0}^{\infty} p_i^m.$$

This series converges for $|p_i| < 1$. For our proposed estimating equations for β, we consider using weight functions of the form

$$w(p_i, M) = \sum_{m=0}^{M} p_i^m, \tag{2.6}$$

for different finite values of M. Note that the Poisson regression estimating equations can be considered the Maclaurin series truncated at $M = 0$ (i.e. constant weights). Higher values of M will more closely approximate the optimal weights associated with the MLE and should yield more efficient estimates.

For any finite value of M, the proposed weight function can be implemented in standard statistical software for generalized linear models that allows user-defined variance functions (e.g. PROC GLIMMIX in SAS); there is negligible increase in computational time for larger values of M. However, some care must be exercised in the choice of value for M; for very large values of M, $w(p_i, M)$ will be very close to $(1 - p_i)^{-1}$, and the resulting estimating equations will be unstable when p_i is close to 1. Recall that convergence problems with ML estimation arise when p_i is close to 1 and the ith observation receives excessively large weight in the estimating equations. We note here that, because $p_i = \mathrm{e}^{x_i'\beta} > 0$, $w(p_i, M)$

will always be positive. Also, for success probabilities close to 1 the weights are bounded, with largest $w(p_i, M) \approx M + 1$, so that convergence problems are less likely unless very large values of M are chosen. In the next section, we examine the ARE for increasing number of terms (M) in the series and make recommendations for the choice of the number of terms to avoid similar finite sample convergence problems as with ML estimation. The challenge is to find some minimum value of M that provides near-optimal weights but essentially bounds the largest weights when p_i is close to 1.

3. ASYMPTOTIC RELATIVE EFFICIENCY

The goal is to find a weight function that approximates the optimal weight function, $w(p_i) = (1 - p_i)^{-1}$, but avoids assigning extremely large weights when p_i is close to 1. Using the truncated Maclaurin series of $(1 - p_i)^{-1}$, $w(p_i, M) = \sum_{m=0}^{M} p_i^m$, we examine the ARE for increasing finite values of M. For this study of ARE, we consider a log RR regression model with a single covariate,

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

where, for simplicity, we let $x_i \times 100$ have a discrete uniform distribution on the set $\{0, 1, 2, \ldots, 100\}$. We let β_1 be negative, so that $\exp(\beta_0)$ is the maximum value for any p_i; in contrast, $\exp(\beta_0 + \beta_1)$ is the minimum value for any p_i. For the study of ARE, we first specify the maximum, $p_{\max}$, and minimum, $p_{\min}$, values for p_i, which in turn fully specifies the parameters $\beta_0 = \log(p_{\max})$ and $\beta_1 = \log(p_{\min}) - \log(p_{\max})$. Specification of the model in this way allows us to explore properties of the estimators as p_i approaches 1. For different choices of values of $p_{\max}$ and $p_{\min}$, we examine the ARE of the estimator of β_1 based on $w(p_i, M)$ for increasing values of M ranging from 0 to 100. Recall that, when $M = 0$, the weights are constant and equivalent to the Poisson regression estimator. Fixing $p_{\max} = 0.99$, we let $p_{\min}$ range from 0.2 to 0.8.

Given a discrete uniform distribution for x_i, and a Bernoulli distribution for Y_i given x_i, the asymptotic variance of β_1 can be obtained from (2.4) by simply considering an artificial sample comprised of one properly weighted observation for each possible realization of (Y_i, x_i). The weights are determined by the respective joint probabilities of the given realizations. Following Rotnitzky and Wypij (1994), the asymptotic variance of β_1 can be ascertained by weighting each contribution to I_0 and I_1 by its respective probability. That is, we take the expectations of I_0 and I_1, the components of C_β in (2.4), by summing all of the possible realizations weighted by their respective probabilities.

A plot of the ARE for increasing values of M is given in Figure 1. The four panels of Figure 1 display the AREs for $p_{\min} = 0.2, 0.4, 0.6, 0.8$, respectively. As $p_{\min}$ increases, the concentration of success probabilities that are close to 1 increases as the median of the probabilities increases from 0.44 when $p_{\min} = 0.2$ to 0.89 when $p_{\min} = 0.8$. In the four panels of Figure 1, the ARE for the Poisson regression estimator ($M = 0$) is in the 60–70% range. This highlights the loss of efficiency associated with the use of constant weights when the probabilities are not small. These results are in close agreement with those reported by Lumley *and others* (2006). As anticipated, the ARE increases monotonically with increasing values of M. AREs of approximately 95–97% are obtained for the proposed estimator when M is between 40 and 60. We note that, for p_i close to 1, $M = 40$ and $M = 60$ bound the maximum weights at approximately 41 and 61, respectively. For larger values of M, there appears to be diminishing returns in terms of increases in efficiency. More importantly, however, larger values of M seem far more likely to produce problems with convergence in finite samples due to excessive weight assigned to observations with p_i close to 1. In the next section, we examine the finite sample performance of the proposed estimator when $M = 20$, $M = 40$, and $M = 60$.

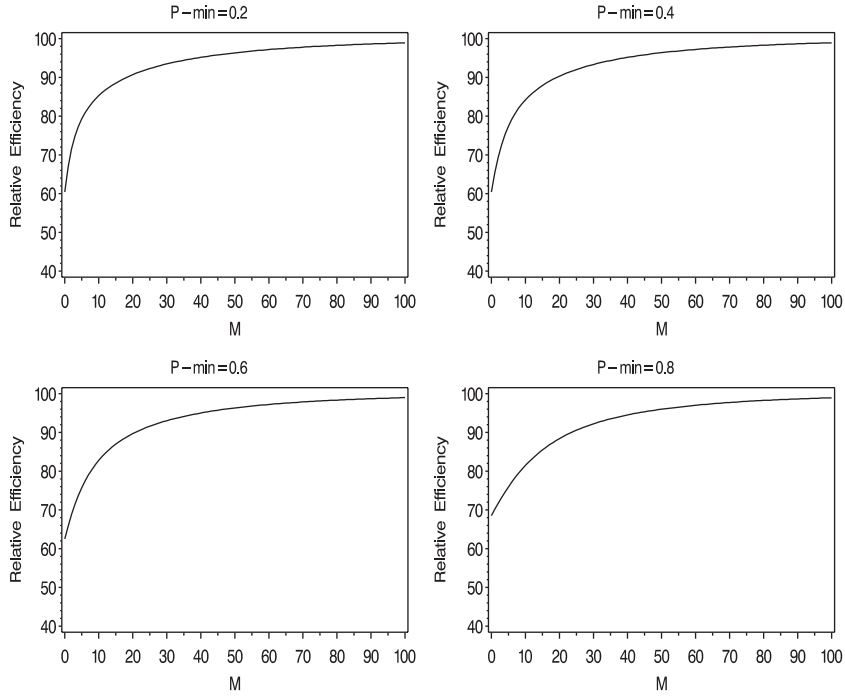

Fig. 1. ARE for increasing number of terms (M) in Maclaurin series expansion of the optimal Bernoulli weight function.

4. Simulation study

In this section, we consider the finite sample properties of the proposed estimator. For the simulations, we used a similar configuration as in the study of ARE presented in Section 3:

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

where, for simplicity, we let x_i have a uniform (0,1) distribution. For the different values of $p_{\max}$ and $p_{\min}$ (or equivalently, β_0 and β_1), we conducted simulations for $n = 100$ with 2500 simulation replications performed for each configuration. We performed simulations fixing $p_{\max} = 0.99$ and letting $p_{\min}$ range from 0.2 to 0.8 in 0.2 unit increments. The simulations were used to compare the MLE (unconstrained), the Poisson regression estimator ($M = 0$), and the proposed estimator based on a Maclaurin series approximation with $M = 20$, $M = 40$, and $M = 60$.

Table 2 presents the relative bias, defined as $100(\hat{\beta}_1 - \beta_1)/\beta_1$, the root mean square error, the coverage probabilities of 95% Wald confidence intervals for β_1, as well as the percentage of simulation replications in which the convergence criterion was met. From Table 2, we see that the percent convergence for ML is less than 50% for all configurations; for ML, we report the results only for those simulations that converged for ML; the latter results can be considered "conditional on the likelihood convergence criterion". In contrast, there were no convergence problems for any of the estimating equations approaches. In terms of bias, standard ML has non-negligible relative bias, and the relative bias increases as $p_{\min}$ increases. The relative bias varies from approximately 8% to 30%. All other approaches have negligible (<5%) relative biases. For the estimating equations approaches, the relative efficiencies can be estimated as the square of the ratio

Table 2. *Simulation results for estimation of* β_1 *for* $\log(p_i) = \beta_0 + \beta_1 x_i$, $x_i \sim \text{Uniform}(0, 1)$, *and* $n = 100$

$p_{\min}$	Method	Percent relative bias of β_1	Root MSE	Coverage probability (%)	Percent converged (%)
0.2	Standard ML	−7.51	3.103	95.6	47.9
	Poisson	2.64	3.767	95.8	100.0
	MS ($M = 20$)	0.97	3.097	95.7	100.0
	MS ($M = 40$)	0.50	3.019	96.1	100.0
	MS ($M = 60$)	0.21	2.985	96.0	100.0
0.4	Standard ML	−12.02	2.278	94.0	48.2
	Poisson	1.60	2.657	94.5	100.0
	MS ($M = 20$)	−0.45	2.183	94.2	100.0
	MS ($M = 40$)	−0.97	2.131	94.4	100.0
	MS ($M = 60$)	−1.33	2.106	94.3	100.0
0.6	Standard ML	−17.53	1.668	93.3	48.0
	Poisson	0.81	1.877	95.0	100.0
	MS ($M = 20$)	−1.50	1.553	95.3	100.0
	MS ($M = 40$)	−2.10	1.505	95.0	100.0
	MS ($M = 60$)	−2.51	1.482	94.9	100.0
0.8	Standard ML	−31.42	1.153	90.5	46.7
	Poisson	1.22	1.244	93.6	100.0
	MS ($M = 20$)	−2.07	1.109	93.2	100.0
	MS ($M = 40$)	−3.02	1.078	93.2	100.0
	MS ($M = 60$)	−3.65	1.063	93.2	100.0

Here, $\beta_0 = \log(p_{\max})$ and $\beta_1 = \log(p_{\min}) - \log(p_{\max})$, where, for a given simulation configuration, $p_{\max} = 0.99$ and $p_{\min}$ are the specified maximum and minimum values of p_i, respectively. MS(M) denotes Maclaurin series expansion of the optimal Bernoulli weight function with M terms.

of the root mean square errors. The relative efficiencies of Poisson regression versus the Maclaurin series with $M = 60$ is between 63% and 73%, increasing from 63% when $p_{\min} = 0.2$ to 73% when $p_{\min} = 0.8$. For these simulations, the relative efficiency of the Maclaurin series with $M = 20$ versus the Maclaurin series with $M = 60$ is above 90% for all configurations (93% when $p_{\min} = 0.2$ and 92% when $p_{\min} = 0.8$). Also, the relative efficiency of the Maclaurin series with $M = 40$ versus the Maclaurin series with $M = 60$ is approximately 97% for all configurations. Interestingly, for these simulation configurations, even use of $M = 20$ terms in the Maclaurin series yields high efficiency.

5. APPLICATION TO STUDY OF PRE-OPERATIVE USE OF BETA BLOCKERS IN PATIENTS WITH COLON CANCER

We apply the proposed methodology to the analysis of pre-operative use of beta blockers (yes/no) among patient undergoing colorectal surgery after diagnosis of colon cancer (Arriaga *and others*, 2009). To examine the relationship between the binary outcome and the five patient-level predictors of interest, we fit the following RR regression model:

$$\log(p_i) = \beta_0 + \beta_1 \text{ASA}_i + \beta_2 \text{Comorbid}_i + \beta_3 \text{Male}_i + \beta_4 \text{White}_i + \beta_5 \text{Age}_i, \tag{5.1}$$

where p_i is the conditional probability that the ith patient receives pre-operative beta blockers; ASA_i is 1 if life-threatening disease and 0 otherwise; Comorbid_i is the number of comorbidities; Male_i is 1 if male

Table 3. *Comparison of* (*log*) *RR regression estimates for the probability of pre-operative use of beta blockers among patients undergoing colorectal surgery*

Effect	Approach	Estimate	SE	Z	p-value
Intercept	Standard ML	−1.186	0.310	−3.82	<0.001
	Poisson	−1.542	0.321	−4.81	<0.001
	MS ($M = 20$)	−1.445	0.262	−5.51	<0.001
	MS ($M = 40$)	−1.445	0.253	−5.71	<0.001
	MS ($M = 60$)	−1.439	0.251	−5.73	<0.001
ASA	Standard ML	−0.009	0.097	−0.10	0.923
	Poisson	−0.017	0.119	−0.15	0.884
	MS ($M = 20$)	−0.007	0.116	−0.06	0.950
	MS ($M = 40$)	−0.007	0.114	−0.07	0.948
	MS ($M = 60$)	−0.008	0.114	−0.07	0.946
Comorbid	Standard ML	0.065	0.044	1.46	0.145
	Poisson	0.080	0.040	2.00	0.047
	MS ($M = 20$)	0.078	0.030	2.60	0.010
	MS ($M = 40$)	0.076	0.028	2.74	0.007
	MS ($M = 60$)	0.074	0.027	2.77	0.006
Male	Standard ML	−0.068	0.098	−0.69	0.488
	Poisson	−0.120	0.107	−1.11	0.266
	MS ($M = 20$)	−0.053	0.088	−0.60	0.552
	MS ($M = 40$)	−0.048	0.080	−0.60	0.551
	MS ($M = 60$)	−0.047	0.076	−0.61	0.541
White	Standard ML	−0.105	0.131	−0.80	0.422
	Poisson	−0.157	0.152	−1.03	0.303
	MS ($M = 20$)	−0.139	0.148	−0.94	0.349
	MS ($M = 40$)	−0.138	0.148	−0.93	0.351
	MS ($M = 60$)	−0.138	0.148	−0.93	0.352
Age	Standard ML	0.111	0.044	2.52	0.013
	Poisson	0.160	0.047	3.41	0.001
	MS ($M = 20$)	0.139	0.037	3.76	<0.001
	MS ($M = 40$)	0.139	0.034	4.08	<0.001
	MS ($M = 60$)	0.139	0.033	4.18	<0.001

MS(M) denotes Maclaurin series expansion of the optimal Bernoulli weight function with M terms.

and 0 if female; White_i is 1 if White race and 0 if otherwise; and Age_i is age in years (although the reported effect of age is multiplied by 10 for easier comparison of results in Table 3).

Table 3 presents the estimates of β obtained using standard ML (as implemented in SAS PROC GENMOD), the Poisson quasi-likelihood approach (McNutt *and others*, 2003; Zou, 2004; Carter *and others*, 2005), and the Maclaurin series approach with $M = 20$, $M = 40$, and $M = 60$. Of note, there were convergence problems with ML but not with any of the other approaches. In particular, for ML SAS PROC GENMOD produced a warning message that "The relative Hessian convergence criterion of 0.0198000361 is greater than the limit of 0.0001. The convergence is questionable".

As conjectured by the study investigators, the results in Table 3 indicate that older patients and patients with more comorbidities are significantly more likely to receive pre-operative beta blockers. From the results presented in Table 3, it is also apparent that the quantitative variable age and the ordinal variable "number of comorbidities" show the largest estimated efficiency gains when comparing the Poisson ($M = 0$) to the Maclaurin series estimators. The relative efficiencies can be estimated as the square of the ratio of the estimated standard errors. For the covariate "number of comorbidities", the estimated relative efficiency of Poisson regression versus the Maclaurin series with $M = 60$ is approximately 46%. For the covariate age, the estimated relative efficiency of Poisson regression versus the Maclaurin series with $M = 60$ is approximately 49%. Thus, in this particular application, Poisson regression appears to be quite inefficient compared with the proposed Maclaurin series approach.

For the covariate "number of comorbidities", the estimated relative efficiency of Maclaurin series with $M = 20$ versus the Maclaurin series with $M = 60$ is greater than 81%; similarly, for the covariate age, it is above 80%. Thus, in this applications, there appears to be discernible gains from using a Maclaurin series with $M = 60$ instead of $M = 20$. However, there is no appreciable difference between Maclaurin series with $M = 40$ and $M = 60$. This reinforces the results from Sections 3 and 4 where it was found that there are diminishing returns in terms of increases in efficiency when M is greater than 40–60.

6. CONCLUSION

To circumvent the usual convergence problems associated with the ML estimator, we propose estimators that approximate the optimal weight function based on the truncated Maclaurin series of $(1 - p_i)^{-1}$. This use of a near-optimal weight function that bounds the largest weights yields estimators with relatively high efficiency that also avoid convergence problems. In our study of asymptotic efficiency, the proposed estimator with weight function based on 40–60 terms from the Maclaurin series was 95–97% efficient relative to the MLE. This compares favorably to the Poisson regression estimator that was found to be only 60–70% efficient. In simulations with samples of size 50 (data not shown) and 100, we found similar gains in relative efficiency and no convergence problems with the proposed estimator based on 40–60 terms. In addition, the proposed estimator, using any finite number of terms, is straightforward to implement in standard statistical software for generalized linear models that allows user-defined variance functions (e.g. PROC GLIMMIX in SAS). Finally, we note that estimators that approximate the optimal weight function based on truncated series expansions may also be useful for other generalized linear models in which the link function does not respect the natural parameter constraints (e.g. linear or linear-expit models for binary data (Kovalchik *and others*, 2013)).

In general, RR regression is most useful when the scientific goal is to estimate the association between an exposure or intervention and a commonly occurring binary outcome, with appropriate adjustment for additional covariates. However, we note that although the estimating equations (2.3) yield consistent estimators of β for any choice of weight function, including the optimal weight function and the approximation to it proposed in this paper, the estimators are not constrained to produce estimated $p_i \leqslant 1$. As a result, RR regression should be avoided altogether when the scientific goal is to make *predictions*; when the goal is prediction, models where the constraints on the probabilities are automatically satisfied (e.g. logistic regression) should be adopted instead.

ACKNOWLEDGMENTS

Conflict of Interest: None declared.

Funding

We are grateful for the support provided by grants MH054693 and CA160679 from the United States National Institutes of Health (NIH).

References

Abramowitz, M. and Stegun, I. A. (1970). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th printing. New York: Dover Publications.

Arriaga, A. F., Lancaster, R. T., Berry, W. R., Regenbogen, S. E., Lipsitz, S. R., Kaafarani, H. M., Elbardissi, A. W., Desai, P., Ferzoco, S. J., Bleday, R. *and others* (2009). The better colectomy project: association of evidence-based best-practice adherence rates to outcomes in colorectal surgery. *Annals of Surgery* **250**, 507–513.

Baumgarten, M., Seliske, P. and Goldberg, M. S. (1989). Warning regarding the use of GLIM macros for the estimation of risk ratios. *American Journal of Epidemiology* **130**, 1065.

Carter, R. E., Lipsitz, S. R. and Tilley, B. C. (2005). Quasi-likelihood estimation for relative risk regression models. *Biostatistics* **6**, 39–44.

Clogg, C. C., Rubin, D. B., Schenker, N., Schultz, B. and Weidman, L. (1991). Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression. *Journal of the American Statistical Association* **86**, 68–78.

Deddens, J. A., Petersen, M. R. and Lei, X. (2003). Estimation of prevalence ratios when PROC GENMOD does not converge. *Proceeding of the 28th Annual SAS Users Group International Conference*. Cary, NC: SAS Institute Inc., pp. 270–278.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1. Berkeley: University of California Press, pp. 221–233.

Jewell, N. P. (2003). *Statistics for Epidemiology*. New York: Chapman & Hall/CRC Press.

Kovalchik, S. A., Varadhan, R., Fetterman, B., Poitras, N. E., Wacholder, S. and Katki, H. A. (2013). A general binomial regression model to estimate standardized risk differences from binary response data. *Statistics in Medicine* **32**, 808–821.

Lu, M. and Tilley, B. C. (2001). Use of odds ratio or relative risk to measure a treatment effect in clinical trials with multiple correlated binary outcomes: data from the NINDS t-PA stroke trial. *Statistics in Medicine* **20**, 1891–1901.

Lumley, T., Kronmal, R. and Ma, S. (2006). Relative risk regression in medical research: models, contrasts, estimators, and algorithms. *UW Biostatistics Working Paper Series*. Working Paper 293. http://biostats.bepress.com/uwbiostat/paper293.

McNutt, L. A., Wu, C., Xue, X. and Hafner, J. P. (2003). Estimating relative risk in cohort studies and clinical trials of common events. *American Journal of Epidemiology* **157**, 940–943.

Owens, W. D., Felts, J. A. and Spitznagel, Jr, E. L. (1978). ASA physical status classifications: a study of consistency of ratings. *Anesthesiology* **49**, 239–243.

Poldermans, D., Boersma, E., Bax, J. J., Thomson, I. R., van de Ven, L. L., Blankensteijn, J. D., Baars, H. F., Yo, T. I., Trocino, G., Vigna, C., Roelandt, J. R. and van Urk, H. (1999). The effect of bisoprolol on perioperative mortality and myocardial infarction in high-risk patients undergoing vascular surgery. *The New England Journal of Medicine* **341**, 1789–1794.

Rotnitzky, A. (2009). Inverse probability weighted methods. In: Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (editors), *Longitudinal Data Analysis*. Boca Raton: Chapman & Hall/CRC Press, pp. 453–476.

ROTNITZKY, A. AND WYPIJ, D. (1994). A note on the bias of estimators with missing data. *Biometrics* **50**, 1163–1170.

ROYALL, R. M. (1986). Model robust confidence intervals using maximum likelihood estimators. *International Statistical Review* **54**, 221–226.

TRAISSAC, P., MARTIN-PREVEL, Y., DELPEUCH, F. AND MAIRE, B. (1999). Regression logistique vs autres modeles lineaires generalises pour l'estimation de rapports de prevalences. *Rev Epidemiol Sante Publique* **47**, 593–604.

WACHOLDER, S. (1986). Binomial regression in GLIM: estimating risk ratios and risk differences. *American Journal of Epidemiology* **123**, 174–184.

WEDDERBURN, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the gauss newton method. *Biometrika* **61**, 439–447.

WHITE, H. (1982). Maximum likelihood estimation under mis-specified models. *Econometrica* **50**, 1–26.

ZHANG, J. AND YU, K. F. (1998). What's the relative risk?: a method of correcting the odds ratio in cohort studies of common outcomes. *Journal of the American Medical Association* **280**, 1690–1691.

ZOU, G. (2004). A modified Poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology* **159**, 702–706.

[*Received July 18, 2013; revised February 27, 2014; accepted for publication March 4, 2014*]