Quasi-likelihood ratio tests and the Bartlett-type correction for improved inferences of the modified Poisson and least-squares regressions for binary outcomes

Hisashi Noma, PhD*
Department of Interdisciplinary Statistical Mathematics, The Institute of Statistical Mathematics, Tokyo, Japan
The Graduate Institute for Advanced Studies, The Graduate University for Advanced Studies (SOKENDAI), Tokyo, Japan
ORCID: http://orcid.org/0000-0002-2520-9949

Hiroshi Sunada, PhD
Advanced Medicine, Innovation and Clinical Research Center, Tottori University Hospital, Tottori, Japan

Masahiko Gosho, PhD
Department of Biostatistics, Institute of Medicine, University of Tsukuba, Tsukuba, Japan

*Corresponding author: Hisashi Noma
Department of Interdisciplinary Statistical Mathematics
The Institute of Statistical Mathematics
10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan
TEL: +81-50-5533-8440
e-mail: noma@ism.ac.jp

Abstract

Logistic regression has been a standard multivariate analysis method for binary outcomes in clinical and epidemiological studies; however, the odds ratios cannot be interpreted as effect measures directly. The modified Poisson and least-squares regressions are alternative effective methods to provide risk ratio and risk difference estimates. However, their ordinary Wald-type inference methods using the sandwich variance estimator seriously underestimate the statistical errors under small or moderate sample settings. In this article, we develop alternative likelihood-ratio-type inference methods for these regression analyses based on Wedderburn's quasi-likelihood theory. A remarkable advantage of the proposed methods is that we have correct information for the true models (i.e., the binomial log-linear and linear models). Using this modeling information, we develop an effective parametric bootstrap algorithm for accurate inferences. In particular, we propose the Bartlett-type mean calibration approach and bootstrap test-based approach for the quasi-likelihood ratio statistic. In addition, we propose another computationally efficient modified approximate quasi-likelihood ratio statistic whose large sample distribution can be approximated by the χ^2 distribution and its bootstrap inference method. In numerical studies by simulations, the new bootstrap-based methods outperformed the current standard Wald-type confidence interval. We applied these methods to a clinical study of epilepsy.

Key words: generalized linear model, estimating equation, quasi-likelihood, bootstrap, separation problem.

1. Introduction

Logistic regression has been widely used as an effective multivariate analysis method for analyzing binary outcome data. However, the regression coefficients can only be interpreted as log-transformed odds ratios. An odds ratio cannot be interpreted as an effect measure by itself and can only be interpreted as an approximation of risk ratio when the frequency of events is small [1,2]. Thus, the use of the risk ratio and risk difference is alternatively recommended in recent guidelines; e.g., the statistical reporting guidelines for *The New England Journal of Medicine* recommend avoiding using the odds ratio when reporting the results of clinical trials.

To circumvent the use of logistic regression, researchers have considered other binomial regression models with log or identity links [3]. However, the values of the regression functions can exceed the range [0, 1] for these models and the maximum likelihood (ML) estimates cannot often be defined in practice [4,5]. The modified Poisson and least-squares (Gaussian) regression analyses [6,7] have been proposed as alternative quasi-likelihood inference methods for these binomial regression analyses with effectively circumventing the computational difficulties. Although the procedures are simple, i.e., formally fitting the ordinary Poisson and least-squares regression models to binary outcome data, consistent estimators of risk ratios and risk differences are obtained. Founded on the quasi-likelihood estimating equation theory of the generalized linear model (GLM; [8-10], the estimating function is assured to be unbiased even if the distributional assumption is violated, as long as the regression function model is correct.

For inferences of the regression coefficients, the variance estimators should be corrected to the sandwich variance estimator [11] because the distributional assumptions are incorrect. One relevant issue with the sandwich variance estimator is serious bias under small or moderate sample settings [12,13]. Also, the quasi-ML estimators of the

regression coefficients have biases under small sample settings, as is well known for the logistic regression [14,15]. Because of these biases, the Wald-type confidence intervals for both of these methods can seriously underestimate the actual statistical errors under small or moderate sample settings. These properties are especially serious under small and sparse data settings, which is known as the "separation" condition—a situation where the outcome variable separates a predictor variable completely. This situation can occur even for large datasets in cases of rare events or influential covariates.

In this article, to address these issues, we propose alternative likelihood-ratio-type inference methods for the modified Poisson and least-squares regression analyses. For the general GLM theory, inference methods based on the likelihood ratio statistic or profile likelihood have been effective alternatives for the simple Wald-type inference methods [16,17]. However, these methods are founded on the ML-based theory and cannot be simply adapted to the quasi-ML inference methods. We first show that the quasi-likelihood ratio statistics of these two regression models cannot be approximated by the χ^2 distribution through asymptotic expansion. We then discuss how to use these statistics in the inferences as the ML-based theory. As a new method, we propose a mean calibration approach of the quasi-likelihood ratio statistics to adjust their scale mis-specifications and improve the χ^2 approximations via the Bartlett-type correction [18,19]. We also propose applying a bootstrap approach to directly estimate the sample distributions of the quasi-likelihood ratio statistics. We then provide additional simply calculable modified approximate quasi-likelihood ratio statistics that follow the χ^2 distribution asymptotically. They can be used as alternative tools for likelihood ratio statistics under large sample settings and are calculable via simple matrix computations. In addition, for small or moderate sample inferences, we propose applying the bootstrap approach to the modified approximate quasi-likelihood ratio statistics. We discuss

adapting all of these methods for calculating accurate confidence intervals. Numerical evaluations via simulation studies showed that the coverage properties of the bootstrap methods were favorable compared with those of the ordinary Wald-type confidence interval. In addition, we demonstrate the application of the proposed methods to a clinical study of epilepsy. We have developed an R package, `QLRM` (https://github.com/nomahi/QLRM), that can implement the proposed methods via simple commands.

2. Modified Poisson and least-squares regressions

Consider a cohort study consisting of n participants with binary outcome $Y_1, \ldots, Y_n$ (= 1: event occurred, = 0: not occurred) and covariates $\boldsymbol{x}_i = \left(x_{i1}, x_{i2}, \ldots, x_{ip}\right)^T$. Although the binomial regressions with log-link and identity-link functions have been conventionally discussed for multivariate analyses of risk ratio and risk difference, they involve serious computational difficulties because the values of the regression functions do not fall within [0, 1] [4,5]. The modified Poisson and least-squares regressions were proposed by Zou [6] and Cheung [7], respectively, as alternative effective methods for these multivariate analyses. Zou [6] and Cheung [7] proposed formally fitting the Poisson and least-squares regression models, respectively, to the binary outcome data,

$$\log(\mathrm{E}[Y_i|\boldsymbol{x}_i]) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$$

$$\mathrm{E}[Y_i|\boldsymbol{x}_i] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$$

The resultant quasi-ML estimators $\widehat{\boldsymbol{\beta}}$ of the regression coefficients $\boldsymbol{\beta} = \left(\beta_0, \beta_1, \ldots, \beta_p\right)^T$ become consistent estimators of the log-transformed risk ratios and risk differences on the target population.

The validity of these estimating methods is founded on the estimating equation theory of GLM [10],

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{D}_i^T V_i^{-1}(Y_i - \mu_i) = 0$$

Where μ_i is the mean function (= exp ($\boldsymbol{\beta}^T \boldsymbol{x}_i$) for the Poisson model and $= \boldsymbol{\beta}^T \boldsymbol{x}_i$ for the Gaussian model) and $\boldsymbol{D}_i = \partial \mu_i / \partial \boldsymbol{\beta}$; in addition, $V_i = v(\mu_i)$ is the variance function of the outcome variable (= μ_i for the Poisson model and $= 1$ for the Gaussian model) $(i = 1, \ldots, n)$. The functional form of $U(\boldsymbol{\beta})$ indicates that, as long as the functional forms of the regression functions are correctly specified, the estimating functions are unbiased even if the distribution forms are mis-specified. In particular, for the modified least-squares regression, the Gauss–Markov theorem [20,21] is fulfilled and the quasi-ML estimator becomes the best linear unbiased estimator (BLUE); $\widehat{\boldsymbol{\beta}}$ is an unbiased and the most precise estimator for the regression coefficients of the binomial regression function. The standard errors of $\widehat{\boldsymbol{\beta}}$ are consistently estimated by the sandwich variance estimator [11].

3. Quasi-likelihood ratio statistic and the Bartlett-type correction

3.1 Asymptotic distribution of the quasi-likelihood ratio statistic

For the modified Poisson and least-squares regressions, the Wald-type confidence interval using the standard sandwich variance estimator [11] has been widely adopted. However, for conventional logistic regression, the profile likelihood approach based on the likelihood ratio statistic has been used as alternative effective option [16,17]. We consider a test for a composite null hypothesis H_0: $\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t,\text{null}} = \left(\beta_{1,\text{null}}, \ldots, \beta_{q,\text{null}}\right)^T$ vs. H_1: $\boldsymbol{\beta}_t \neq \boldsymbol{\beta}_{t,\text{null}}$; without loss of generality, we consider the test of regression coefficients of first q variables ($q < p$) and denote the complement as $\boldsymbol{\beta}_c = \left(\beta_0, \beta_{q+1}, \ldots, \beta_p\right)^T$. We also denote the constrained quasi-ML estimate of $\boldsymbol{\beta}_c$ under H_0 as $\widetilde{\boldsymbol{\beta}}_c = \left(\tilde{\beta}_0, \tilde{\beta}_{q+1}, \ldots, \tilde{\beta}_p\right)^T$ and denote $\widetilde{\boldsymbol{\beta}} = \left(\tilde{\beta}_0, \beta_{1,\text{null}}, \ldots, \beta_{q,\text{null}}, \tilde{\beta}_{q+1}, \ldots, \tilde{\beta}_p\right)^T$; the

constrained quasi-ML estimates can be computed similarly by setting $\beta_{1,\text{null}}x_{i1} + \cdots + \beta_{q,\text{null}}x_{iq}$ as the offset. The quasi-loglikelihood function $\ell(\boldsymbol{\beta})$ is then expressed via the asymptotic expansion around $\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}$,

$$\ell(\boldsymbol{\beta}) = \ell(\widetilde{\boldsymbol{\beta}}) + U(\widetilde{\boldsymbol{\beta}})(\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}) - \frac{1}{2}(\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}})^T I(\widetilde{\boldsymbol{\beta}})(\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}})$$

where $I(\boldsymbol{\beta}) = -E[\partial U(\boldsymbol{\beta})/\partial\boldsymbol{\beta}]$ is the Fisher information matrix ($= \boldsymbol{X}\boldsymbol{R}_1\boldsymbol{X}^T$ for the modified Poisson regression and $= \boldsymbol{X}\boldsymbol{X}^T$ for the modified least-squares regression; $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$; and $\boldsymbol{R}_1 = \text{diag}\{\mu_1, \ldots, \mu_n\}$). When $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$, the quasi-loglikelihood ratio statistic is approximated as

$$-2\{\ell(\widehat{\boldsymbol{\beta}}) - \ell(\widetilde{\boldsymbol{\beta}})\} = (\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})^T I(\widetilde{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}) \qquad (*)$$

Therefore, the quasi-likelihood statistic is approximated by the standardized multivariate normal quantity $\widehat{\boldsymbol{\beta}}$. When the quasi-ML estimator $\widehat{\boldsymbol{\beta}}$ achieves the Cramer–Rao's bound asymptotically, the asymptotic distribution can be approximated by the χ^2 distribution. However, the variance function is mis-specified for these estimating methods; the asymptotic distribution of $(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})$ corresponds to MVN($\mathbf{0}$, $I(\widetilde{\boldsymbol{\beta}})^{-1}J(\widetilde{\boldsymbol{\beta}})I(\widetilde{\boldsymbol{\beta}})^{-1}$), where $J(\boldsymbol{\beta}) = E[U(\boldsymbol{\beta})U(\boldsymbol{\beta})^T]$. Thus, the quantity on the right-hand side does not follow a χ^2 distribution even under large sample settings and we cannot use the profile likelihood approach for these analysis methods simply as the conventional ML-based theory.

3.2 Mean calibration by bootstrapping approach

The asymptotic distribution of the quasi-likelihood ratio statistic

$$T(\boldsymbol{\beta}_{t,\text{null}}) = -2\{\ell(\widehat{\boldsymbol{\beta}}) - \ell(\widetilde{\boldsymbol{\beta}})\}$$

cannot be approximated by the χ^2 distribution, and the preceding result indicates that the standardization term $I(\widetilde{\boldsymbol{\beta}})$ of right hand of (*) is inadequate for the χ^2

approximation, i.e., only the dispersion adjustment is inadequate. One effective approach to constructing a valid statistical test using the quasi-likelihood ratio statistic is to adjust the dispersion correctly, and a simple method to carry out the adjustment is to multiply the adjustment term by the quasi-likelihood ratio statistic. In addition, a conventional approach to improve the likelihood ratio test via higher-order asymptotic theory is the Bartlett correction [18,19] that improves the χ^2 approximation via multiplication by an adjustment term: the inverse of an estimate of the mean of sample distribution of $T(\boldsymbol{\beta}_{t,\text{null}})$, $E[T(\boldsymbol{\beta}_{t,\text{null}})]$. The former dispersion adjustment is also achieved by the mean calibration because the χ^2 variable is only mis-standardized by the inadequate standardization term. Thus, both adjustments are realized by calibrating $T(\boldsymbol{\beta}_{t,\text{null}})$ against the estimated mean of the sample distribution. We here propose adjusting both of these factors simultaneously by dividing the quasi-likelihood ratio statistic by the bootstrap estimate of the mean of the sample distribution,

$$T_c(\boldsymbol{\beta}_{t,\text{null}}) = \frac{T(\boldsymbol{\beta}_{t,\text{null}})}{C}$$

where C is the bootstrap estimate of $E[T(\boldsymbol{\beta}_{t,\text{null}})]$. In applying the bootstrapping approach, it should be noted that a relevant feature of these estimating methods is that the correct distributional assumptions about the target population are known; the binomial log-linear and linear regression models. This issue is a distinguishable characteristic of these methods from other semiparametric regression models, e.g., the generalized estimating equation (GEE), in which the second moment assumption is substantially unknown and working assumptions are formally adapted [22]. In moderate sample analyses, the parametric assumptions can gain the accuracy of bootstrap inferences and we adopt the parametric bootstrap approach using the correct model information here. We propose performing the bootstrap resampling from the “correct” binomial regression

models by substituting the regression coefficients $\boldsymbol{\beta}$ for the null value $\widetilde{\boldsymbol{\beta}}$. The bootstrap algorithm to estimate the adjustment term C is described as follows:

Algorithm 1 (bootstrap estimation for the calibration term C)

1. For the modified Poisson and least-squares regression models, compute the constrained quasi-ML estimates $\widetilde{\boldsymbol{\beta}}_c$ under H_0: $\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t,\text{null}}$.

2. Resample $Y_1^{(b)}, \ldots, Y_n^{(b)}$ from the binomial regression models with log or identity links whose regression coefficients $\boldsymbol{\beta}$ are fixed to $\widetilde{\boldsymbol{\beta}}$,

$$\Pr(Y_i = 1|\boldsymbol{x}_i) = \exp(\tilde{\beta}_0 + \beta_{1,\text{null}}x_{i1} + \cdots + \beta_{q,\text{null}}x_{iq} + \tilde{\beta}_{q+1}x_{i\,q+1} + \cdots + \tilde{\beta}_p x_{ip})$$

$$\Pr(Y_i = 1|\boldsymbol{x}_i) = \tilde{\beta}_0 + \beta_{1,\text{null}}x_{i1} + \cdots + \beta_{q,\text{null}}x_{iq} + \tilde{\beta}_{q+1}x_{i\,q+1} + \cdots + \tilde{\beta}_p x_{ip}$$

via parametric bootstraps, B times ($i = 1, \ldots, n;\ b = 1,2, \ldots, B$). The design matrix $\boldsymbol{X}$ is not altered from the original data across the B resampling. Also, if the values of the regression functions on the right-hand sides of the equations exceed the range [0, 1], they should be truncated at 0 or 1.

3. Compute the quasi-likelihood ratio statistic $T^{(b)}(\boldsymbol{\beta}_{t,\text{null}})$ for the bth bootstrap sample $Y_1^{(b)}, \ldots, Y_n^{(b)}$ ($b = 1,2, \ldots, B$).

4. Calculate the empirical mean of $T^{(1)}(\boldsymbol{\beta}_{t,\text{null}}), \ldots, T^{(B)}(\boldsymbol{\beta}_{t,\text{null}})$ as a bootstrap estimate of the adjustment term, i.e., $C = \sum_{b=1}^{B} T^{(b)}(\boldsymbol{\beta}_{t,\text{null}})/B$.

We propose then using the resultant $T_c(\boldsymbol{\beta}_{t,\text{null}})$ as the test statistic and to adopt the ordinary χ^2 distribution with q degrees of freedom as the reference distribution. In addition, when $q = 1$, the $100 \times (1-\alpha)\%$ confidence intervals of β_1 can be constructed by the sets of null values that satisfy

$$T_c(\beta_{1,\text{null}}) \leq \chi^2_{1,1-\alpha}$$

where $\chi^2_{1,1-\alpha}$ is the upper αth percentile of the χ^2 distribution with 1 degree of

freedom. The confidence limits can be calculated by adequate numerical methods (e.g., the bisectional methods; [23]). The accuracy of the test and confidence interval is expected to be improved by the higher-order approximation based on the Bartlett correction theory [18,19]. The adjusted quasi-likelihood ratio test can be generally used like the ordinary likelihood ratio test for the ML-based approaches of GLM [24] involving the profile-likelihood analyses [25]. Notably, to ensure the accuracy of the bootstrap mean estimation, the number of bootstrap resampling B should be sufficiently large (usually at least 200; [26]).

3.3 Bootstrap inferences via direct estimation of the sample distribution of $T(\boldsymbol{\beta}_{t,null})$

Another effective approach to construct statistical tests or confidence intervals is the direct estimation of the sample distribution via bootstrap. In this approach, we adopt the quasi-likelihood ratio statistic $T(\boldsymbol{\beta}_{t,\text{null}})$ as the test statistic and use its bootstrap distribution as the reference distribution. The bootstrap algorithm is as follows:

Algorithm 2 (bootstrap estimation for the sample distribution of $T(\boldsymbol{\beta}_{t,\text{null}})$*)*

1. Implement procedures 1–3 of Algorithm 1 and generate the bootstrap samples $T^{(1)}(\boldsymbol{\beta}_{t,\text{null}}), \ldots, T^{(B)}(\boldsymbol{\beta}_{t,\text{null}})$.

2. Calculate the empirical distribution function of $T^{(1)}(\boldsymbol{\beta}_{t,\text{null}}), \ldots, T^{(B)}(\boldsymbol{\beta}_{t,\text{null}})$, specifically, $\bar{F}_T(t)$, which corresponds to the bootstrap estimate of the sample distribution of $T(\boldsymbol{\beta}_{t,\text{null}})$.

To control the accuracy of the Monte Carlo estimation for the tail area of the null distribution, the number of replications should be sufficiently large (usually at least 1000; [26]). The bootstrap-based test simply uses $\bar{F}_T(t)$ as the reference distribution instead of

the χ^2 distribution. Also, when $q = 1$, the corresponding $100 \times (1-\alpha)\%$ confidence interval of β_1 can be constructed by the sets of $\beta_{1,\text{null}}$ that fulfill

$$T(\beta_{1,\text{null}}) \leq \bar{F}_T(1-\alpha)$$

The confidence limits can also be calculated by adequate numerical methods (e.g., the bisectional methods; [23]). The parametric bootstrap approach also effectively uses the distributional information of the correct models; thus, the approximation of the sample distribution is expected to be improved. The actual performances are demonstrated in the simulation studies presented in Section 4.

3.4 Scale adjustment of the asymptotic approximation

For the asymptotic approximation of the quasi-likelihood ratio statistic (*), the quantity on the right-hand side mis-specifies the scaling factor. However, this quantity is similar to the ordinary Wald statistic and is easy to calculate by matrix computation in practice. Also, it involves the quasi-ML estimates under both the null and alternative hypotheses and, with some modifications, might be used as an approximate quantity for the scale-adjusted quasi-likelihood statistic. We then propose the following statistic as the approximate scale-corrected quasi-likelihood ratio statistic, i.e.,

$$W(\boldsymbol{\beta}_{t,\text{null}}) = \left(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}\right)^T \left\{ I\left(\widetilde{\boldsymbol{\beta}}\right)^{-1} J\left(\widetilde{\boldsymbol{\beta}}\right) I\left(\widetilde{\boldsymbol{\beta}}\right)^{-1} \right\}^{-1} \left(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}\right)$$

This statistic can be interpreted as an approximated quantity of the mean calibrated quasi-likelihood statistic described in Section 3.2. Note that the center portion of the sandwich estimator $J(\boldsymbol{\beta})$ can be estimated using the "correct" model information, which is the distinguishing feature of these quasi-likelihood regression analyses, as previously noted, i.e., the expectations can be substituted for those of the binomial regression models. The concrete form is expressed as

$$J(\boldsymbol{\beta}) = \sum_{i=1}^{n} V[Y_i]\boldsymbol{x}_i\boldsymbol{x}_i^T = \sum_{i=1}^{n} \mu_i(1-\mu_i)\boldsymbol{x}_i\boldsymbol{x}_i^T = \boldsymbol{X}\boldsymbol{R}_2\boldsymbol{X}^T$$

where $\boldsymbol{R}_2 = \text{diag}\{\mu_1(1-\mu_1), \ldots, \mu_n(1-\mu_n)\}$. Although the definitions of the mean function μ_i differ between the Poisson and Gaussian models, the functional forms are the same. Note that $\mu_1, \ldots, \mu_n$ should be truncated on [0, 1] on $\boldsymbol{R}_2$ because the individual variance functions should not be negative values.

The scale-adjusted approximate statistic $W(\boldsymbol{\beta}_{t,\text{null}})$ can be computed easily by standard statistical software and can be directly used like the conventional likelihood statistic for GLM [24]. If $W(\boldsymbol{\beta}_{t,\text{null}})$ is used as the test statistic, the asymptotic distribution is approximated by the ordinary χ^2 distribution with q degrees of freedom and can be used like the conventional χ^2-tests. Also, when $q = 1$, the $100 \times (1-\alpha)\%$ confidence intervals of β_1 can be constructed by the sets of null values that satisfy

$$W(\beta_{1,\text{null}}) \leq \chi^2_{1,1-\alpha}$$

where $\chi^2_{1,1-\alpha}$ is the upper αth percentile of the χ^2 distribution with one degree of freedom. The confidence limits can also be calculated by adequate numerical methods (e.g., the bisectional methods; [23]).

In addition, the χ^2 approximation can also be violated under small or moderate sample settings for $W(\boldsymbol{\beta}_{t,\text{null}})$ because this approximation depends on the central limit theorem. To address this problem, we can also adopt the bootstrap approach described in Section 3.3 for the inferences based on $W(\boldsymbol{\beta}_{t,\text{null}})$. The bootstrap algorithm is provided as follows:

Algorithm 3 (bootstrap estimation for the sample distribution of $W(\boldsymbol{\beta}_{t,\text{null}})$)

1. Implement procedures 1–3 of Algorithm 1 and generate the bootstrap samples $W^{(1)}(\boldsymbol{\beta}_{t,\text{null}}), \ldots, W^{(B)}(\boldsymbol{\beta}_{t,\text{null}})$ by calculating $W(\boldsymbol{\beta}_{t,\text{null}})$ for the bth bootstrap sample

$Y_1^{(b)}, \ldots, Y_n^{(b)}$ $(b = 1,2, \ldots, B)$.

2. Calculate the empirical distribution function of $W^{(1)}(\boldsymbol{\beta}_{t,\text{null}}), \ldots, W^{(B)}(\boldsymbol{\beta}_{t,\text{null}})$, specifically, $\bar{F}_W(t)$, which corresponds to the bootstrap estimate of the sample distribution of $W(\boldsymbol{\beta}_{t,\text{null}})$.

To control the Monte Carlo error in the bootstrap estimation, the number of replications should also be sufficiently large (usually at least 1000; [26]). The bootstrap-based test can be constructed using $\bar{F}_W(t)$ as the reference distribution instead of the χ^2 distribution. Also, when $q = 1$, the $100 \times (1 - \alpha)\%$ confidence interval of β_1 can be constructed by the sets of $\beta_{1,\text{null}}$ that fulfill

$$W(\beta_{1,\text{null}}) \leq \bar{F}_W(1 - \alpha)$$

The confidence limits can also be calculated by adequate numerical methods, e.g., the bisectional methods [23]. The bootstrap approach effectively adapts the higher-order asymptotic for the statistical inferences, and the accuracy of the inferences is expected to be improved. The performances under actual data analyses are demonstrated in the simulation studies presented in Section 4.

An R package, `QLRM`, for implementing the proposed methods via simple commands is available at https://github.com/nomahi/QLRM.

4. Simulations

To assess the performances of the proposed methods under practical settings, we carried out simulation experiments. We generated the datasets from the binomial regression models with log and identity links, and their parameter settings mimicked the clinical study of epilepsy discussed in Section 5. In the regression functions, four variables were modelled: x_{i1} was the main treatment/exposure variable that followed a Bernoulli

distribution with probability 0.20 or 0.10 ($= P_{\text{treat}}$); x_{i2} was a confounding variable that followed a Bernoulli distribution with probability 0.773 and had a correlation with x_{i1} measured by the odds ratio (OR) $\Pr(x_{i1} = 1)\Pr(x_{i2} = 0)/\Pr(x_{i1} = 0)\Pr(x_{i2} = 1)$ = 25, 15, and 5; x_{i3} followed a Bernoulli distribution with probability 0.455; and x_{i4} followed N(29.0, 7.37). On the basis of the regression functions, the outcome variable Y_i was generated from a Bernoulli distribution with probability

$$\Pr(Y_i = 1) = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4})$$

$$\Pr(Y_i = 1) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}$$

The intercept β_0 was set by controlling the overall event rate of the cohort; the event rate was varied on 0.40 and 0.20 ($= P_{\text{event}}$). The other regression coefficients were set as $(\beta_1, \beta_2, \beta_3, \beta_4) = (0.205, -0.271, 0.000, 0.153)$ for the log-linear model and $(\beta_1, \beta_2, \beta_3, \beta_4) = (0.116, -0.041, 0.0037, 0.023)$ for the linear model. The sample size was set as $n = 40, 60, 80$. We implemented 5000 simulations for each scenario, both for the log-linear and linear regression models.

For performance measurements, we assessed the coverage probabilities of 95% confidence intervals for the regression coefficient of the main treatment/exposure variable. We applied the four proposed methods in Section 3 for analyses of individual datasets and calculated 95% confidence intervals by (1) the corrected quasi-likelihood ratio (CQLR) by mean calibration, (2) the bootstrap approach based on the quasi-likelihood ratio (QLR), (3) the modified approximate quasi-likelihood ratio (MAQLR), (4) the bootstrap approach based on the MAQLR statistic. For the bootstrap methods, we consistently performed 2000 bootstrap resamplings. For comparisons, we also applied the current standard ordinary Wald-type confidence interval. The coverage probabilities for the true risk ratios and risk differences were evaluated.

The results of the simulations are presented in Tables 1 and 2. The empirical coverage probabilities of 95% confidence intervals for the 5000 simulations are summarized for the five methods. For the modified Poisson regression, the Wald 95% confidence intervals generally showed undercoverage under moderate sample settings, especially for scenarios where the separations or quasi-separations are likely to occur; that is, where P_{treat} and P_{event} are small and the OR is large. These results were caused by the biases for both the regression coefficient and the robust standard error estimates under these conditions. Also, the MAQLR confidence interval without bootstrap showed better small-sample performances compared with the ordinary Wald confidence interval under certain scenarios but also showed worse performances under other scenarios. This difference might be attributable to this method also depending on the large-sample χ^2 approximation without higher-order approximations. In addition, the other three bootstrap-based methods retained the coverage probabilities of approximately the nominal level (95%) for almost all of the scenarios. Even for small sample settings ($n = 40$), the coverage probabilities were valued at approximately the nominal level (95%) consistently. For the CQLR confidence interval, although the χ^2 approximation was used, the coverage performances were clearly improved compared with the naïve methods because the Bartlett-type correction was applied. In addition, the bootstrap approaches for estimating the sample distribution improved the approximations and the coverage performances were clearly improved.

For the modified least-squares regression, the overall features of the results were similar to those for the modified Poisson regression cases. Under small or moderate sample settings, the Wald-type 95% confidence intervals showed undercoverage in general. Note that the quasi-ML estimates obtained by the modified least-squares regression correspond to the BLUE obtained by the Gauss–Markov theorem; thus, no

biases were observed for the coefficient estimator. However, the robust variance estimator was seriously biased under small or moderate sample settings. The degrees of undercoverage also became serious if P_{treat} and P_{event} were small and the OR was large. Similar trends were observed for the MAQLR confidence interval by the ordinary χ^2 approximation. However, the bootstrap-based approaches provided valid confidence intervals with favorable coverage properties consistently. Even for the small sample settings ($n = 40$), the coverage probabilities were consistently close to the nominal level (95%). The coverage probabilities were relatively larger than those of the modified Poisson regression cases, which might be caused by the regression coefficient estimator being unbiased for the modified least-squares regression even under small sample settings. For larger sample settings, the accuracy of the confidence intervals based on the proposed three methods would be retained.

5. Applications

We applied the proposed methods to the epilepsy clinical study reported by Arai et al. [27]. This study was performed using electric medical records to determine what factors are associated with the employment statuses of patients with a history of childhood-onset drug-resistant epilepsy. Here, we analyzed 44 patients who lived in Tottori prefecture in Japan. The outcome was employment status (1 = non-employment, 0 = employment; the number of events was 11), and we analyzed this retrospective cohort data using the modified Poisson and least-squares regressions. Four explanatory variables were modeled in the regression models: age at follow up, gender, mood disorder symptoms, and graduating from a school for special needs education. Note the last two variables were highly correlated, and both of these variables were significantly associated with the outcome through univariate analyses [27]. For the bootstrap-based methods, we

consistently performed 2000 resamplings.

The results are presented in Table 3. The quasi-ML estimates of risk ratios and risk differences and their 95% confidence intervals using the ordinary Wald-type method and the proposed four methods are presented. A remarkable characteristic of the proposed methods is that the confidence intervals are asymmetric around the quasi-ML estimates in general. Also, in some cases, locations of the confidence intervals differ substantially from the Wald-type confidence intervals. These features are generally known for the likelihood ratio-based confidence intervals [28]. The widths of the confidence intervals obtained by the bootstrap methods were generally larger than the Wald confidence intervals.

For the modified Poisson regression, the 95% Wald-type confidence intervals covered the null value (=1) for all variables. The proposed confidence intervals also all covered 1. For the third covariate, the bootstrap-based confidence intervals were narrower than that of Wald confidence interval. Such phenomena have sometimes been verified in our experience. However, the coverage probabilities should be the nominal level on average, which means that these confidence intervals can provide more precise interval estimates while retaining their coverage validities. For the bootstrap confidence interval obtained by the MAQLR for gender, the upper limit of the confidence interval was not identifiable; the bootstrap *P*-values for the upper region of the quasi-ML estimate were all >0.05. Such results can also occur sometimes, but the coverage probabilities should be retained at the nominal level. However, in such cases, other methods should also be attempted. In addition, for the modified least-squares regression, neither the Wald-type confidence interval of the third covariate nor the confidence intervals obtained by the proposed methods involved the null value (=0). However, the confidence intervals obtained by the bootstrap-based approaches were wider and slightly vague effects were

indicated. Also, for the fourth covariate, the CQLR and bootstrapping QLR confidence intervals were much wider than the Wald confidence interval; however, the MAQLR confidence interval did not substantially differ from the Wald confidence interval. The simulation studies presented in Section 4 showed that the ordinary Wald confidence interval possibly underestimates the statistical errors and that the proposed bootstrap methods provide more accurate interval estimates in general. The improved methods would provide more precise evidence in these clinical studies.

6. Discussion

In clinical and epidemiological studies, odds ratios can lead to misleading interpretations, and the applications of logistic regression analyses have been seriously limited [1-3]. The modified Poisson and least-squares regressions provide effective solutions to this relevant issue without introducing computational difficulties and have been increasingly used in recent clinical and epidemiological studies. However, in this article, we explicitly showed that the ordinary Wald-type confidence intervals obtained in these methods could seriously underestimate the statistical errors, especially under separation or quasi-separation settings. This inadequacy is clearly resolved by the proposed new methods. Given the enormous effects of misleading evidence on clinical practice, public health, and policy making, accurate inference methods should be adopted in clinical and epidemiological studies. On the basis of the numerical evidence presented in this study, we expect that our new methods will be recommended in practical applications even under large sample settings in future medical studies.

Comparing the four new methods, we find that the MAQLR-based inference could possibly produce erroneous results under moderate sample settings. However, the large sample approximation will be fulfilled under sufficiently large sample settings. Then, the

MAQLR-based method is computationally efficient because it uses the χ^2 approximation and can be used as an alternative method to the conventional likelihood-ratio-based analysis methods of GLM, such as deviance analyses [24]. The other bootstrap-based approaches will certainly provide more accurate results, and recent computational environments enable these methods to be used as realistically applicable methods for large-scale datasets. In the three bootstrap methods, the bootstrap testing-based methods require relatively large resampling numbers to control the accuracy of the Monte Carlo estimation of the quantiles [26]. In addition, the Bartlett-type mean calibration approach only requires a mean estimation and the Monte Carlo error can be controlled with a relatively small number of resamplings (e.g., 200). The performances of the inferences were not substantially different in the numerical studies, and the mean calibration approach might be recommended if the computational effort is a relevant concern.

In future studies, the bias correction of the quasi-ML estimator will be a relevant concern because the Wald-type confidence interval failed to assess statistical errors under most scenarios. This issue is specifically important under separation or quasi-separation settings; although no relevant solutions have been found, a general theory for the ML estimator of GLM has been established. The Firth-type bias corrections are developed by Uno et al. [29], and higher-order bias correction methods [30] will also be applicable; they are further relevant issues. Modifications of the robust variance estimator [12] would be another solution. Although the accuracies achieved with the existing improved robust variance estimators are generally better than those achieved with the ordinary sandwich variance estimator, deterministic conclusions cannot be provided for their relative performances [12] because all the methods are founded on some approximations (e.g., higher-order approximations). Although simulation-based numerical evidence enables

case-by-case performance comparisons, these comparisons would not provide generic conclusions. Nonetheless, the bootstrap-based approaches proposed in this article effectively apply the information of the null hypothesis and adapt the higher-order asymptotics. We speculate that they would be one of the most effective methods among the methods developed thus far. We recommend the proposed methods as accurate and effective alternatives to the Wald-type inference methods in clinical and epidemiological studies.

Acknowledgements

We would like to thank Dr. Y. Arai (Tottori University) for permission to use the valuable data.

Research funding

This study was supported by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (grant numbers: JP23K11931, JP22H03554, JP24K21306, and JP23H03063).

Conflict of interest statement

The authors declare no conflicts of interest regarding this article.

References

1. Greenland S. Interpretation and choice of effect measures in epidemiologic analysis. *Am J Epidemiol*. 1987;125:761-768.
2. Nurminen M. To use or not to use the odds ratio in epidemiologic analyses. *Eur J Epidemiol*. 1995;11:365-371.

3. Rothman KJ, Greenland G, Lash TL. *Modern Epidemiology*. 3rd ed. Lippincott Williams & Wilkins; 2008.
4. McNutt LA, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol*. 2003;157:940-3.
5. Wallenstein S, Bodian C. Epidemiologic programs for computers and calculators. Inferences on odds ratios, relative risks, and risk differences based on standard regression programs. *Am J Epidemiol*. 1987;126:346-55.
6. Zou G. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol*. 2004;159:702-6.
7. Cheung YB. A modified least-squares regression approach to the estimation of risk difference. *Am J Epidemiol*. 2007;166:1337-44.
8. Nelder JA, Wedderburn RWM. Generalized linear models. *J Royal Stat Soc A*. 1972;135:370-384.
9. Godambe VP, Heyde CC. Quasi-likelihood and optimal estimation. *Int Stat Rev*. 1987;55:231-244.
10. Wedderburn RWM. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*. 1974;61:439-447.
11. White H. Maximum likelihood estimation of misspecified models. *Econometrica*. 1982;50:1-25.
12. Gosho M, Ishii R, Noma H, Maruo K. A comparison of bias-adjusted generalized estimating equations for sparse binary data in small-sample longitudinal studies. *Stat Med*. 2023;42:2711-2727.
13. Gosho M, Noma H, Maruo K. Practical review and comparison of modified covariance estimators for linear mixed models in small-sample longitudinal studies with missing data. *Int Stat Rev*. 2021;89:550-572.

14. Albert A, Anderson JA. On the existence of the maximum likelihood estimates in logistic regression models. *Biometrika*. 1984;71:1-10.
15. Zorn C. A solution to separation in binary response models. *Polit Anal*. 2005;13:157-170.
16. Heinze G. A comparative investigation of methods for logistic regression with separated or nearly separated data. *Stat Med*. 2006;25:4216-4226.
17. Lang JB. Score and profile likelihood confidence intervals for contingency table parameters. *Stat Med*. 2008;27:5975-90.
18. Bartlett MS. Properties of sufficiency and statistical tests. *Proc Royal Soc A*. 1937;160:268-282.
19. Cordeiro GM, Cribari-Neto F. *An Introduction to Bartlett Correction and Bias Reduction*. Springer; 2014.
20. Plackett RL. A historical note on the method of least squares. *Biometrika*. 1949;36:458-460.
21. David FN, Neyman J. Extension of the Markoff theorem on least squares. *Stat Res Mem*. 1938;2:105-116.
22. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73:13-22.
23. Burden RL, Faires JD. *Numerical Analysis*. 9th ed ed. Cengage Learning; 2011.
24. Agresti A. *Categorical Data Analysis*. 3rd ed. Wiley; 2013.
25. Venzon DJ, Moolgavkar SH. A method for computing profile likelihood-based confidence intervals. *Appl Statist*. 1988;37:87-94.
26. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. CRC Press; 1994.
27. Arai Y, Okanishi T, Noma H, et al. Prognostic factors for employment outcomes in patients with a history of childhood-onset drug-resistant epilepsy. *Front Pediatr*.

2023;11:1173126.

28. Noma H. Confidence intervals for a random-effects meta-analysis based on Bartlett-type corrections. *Stat Med*. 2011;30:3304-3312.
29. Uno S, Noma H, Gosho M. Firth-type penalized methods of the modified Poisson and least-squares regression analyses for binary outcomes. *Biom J*. 2024;66:e202400004.
30. Cordeiro GM, McCullagh P. Bias correction in generalized linear models. *J Royal Stat Soc B*. 1991;53:629-643.

Table 1. Results of the simulation studies (I): Monte Carlo estimates of coverage probabilities of the 95% confidence intervals.

				Modified Poisson regression					Modified least-squares regression				
P_{event}	P_{treat}	OR	n	Wald	CQLR	QLR (bootstrap)	MAQLR	MAQLR (bootstrap)	Wald	CQLR	QLR (bootstrap)	MAQLR	MAQLR (bootstrap)
0.4	0.2	25	40	0.925	0.941	0.939	0.896	0.944	0.899	0.957	0.957	0.943	0.954
0.4	0.2	25	60	0.957	0.951	0.951	0.923	0.955	0.913	0.948	0.948	0.944	0.947
0.4	0.2	25	80	0.955	0.949	0.949	0.929	0.951	0.921	0.950	0.949	0.943	0.947
0.4	0.2	15	40	0.936	0.948	0.949	0.903	0.951	0.894	0.950	0.950	0.938	0.947
0.4	0.2	15	60	0.945	0.943	0.942	0.917	0.948	0.914	0.947	0.947	0.938	0.945
0.4	0.2	15	80	0.956	0.947	0.946	0.930	0.947	0.927	0.952	0.951	0.946	0.950
0.4	0.2	5	40	0.926	0.946	0.946	0.904	0.951	0.886	0.950	0.949	0.940	0.950
0.4	0.2	5	60	0.946	0.945	0.945	0.918	0.948	0.914	0.950	0.950	0.943	0.949
0.4	0.2	5	80	0.957	0.954	0.953	0.936	0.956	0.924	0.951	0.950	0.945	0.949
0.4	0.1	25	40	0.760	0.948	0.949	0.829	0.959	0.741	0.959	0.956	0.947	0.951
0.4	0.1	25	60	0.867	0.946	0.945	0.889	0.949	0.852	0.958	0.957	0.950	0.954
0.4	0.1	25	80	0.911	0.943	0.943	0.900	0.943	0.888	0.953	0.951	0.946	0.950
0.4	0.1	15	40	0.750	0.949	0.945	0.818	0.953	0.749	0.958	0.955	0.948	0.951
0.4	0.1	15	60	0.866	0.942	0.941	0.885	0.946	0.846	0.958	0.955	0.949	0.952
0.4	0.1	15	80	0.924	0.952	0.951	0.914	0.952	0.886	0.950	0.948	0.945	0.946
0.4	0.1	5	40	0.754	0.945	0.946	0.818	0.951	0.742	0.964	0.960	0.948	0.952
0.4	0.1	5	60	0.865	0.940	0.942	0.879	0.944	0.848	0.958	0.957	0.952	0.955
0.4	0.1	5	80	0.916	0.944	0.945	0.904	0.945	0.885	0.953	0.952	0.947	0.950

Table 2. Results of the simulation studies (II): Monte Carlo estimates of coverage probabilities of the 95% confidence intervals.

				Modified Poisson regression					Modified least-squares regression				
P_{event}	P_{treat}	OR	n	Wald	CQLR	QLR (bootstrap)	MAQLR	MAQLR (bootstrap)	Wald	CQLR	QLR (bootstrap)	MAQLR	MAQLR (bootstrap)
0.2	0.2	25	40	0.789	0.946	0.941	0.783	0.955	0.860	0.973	0.959	0.940	0.958
0.2	0.2	25	60	0.913	0.950	0.950	0.885	0.953	0.906	0.957	0.953	0.940	0.953
0.2	0.2	25	80	0.939	0.945	0.947	0.906	0.950	0.917	0.954	0.953	0.942	0.952
0.2	0.2	15	40	0.806	0.947	0.944	0.795	0.958	0.846	0.968	0.953	0.934	0.955
0.2	0.2	15	60	0.907	0.946	0.945	0.877	0.952	0.904	0.957	0.955	0.942	0.956
0.2	0.2	15	80	0.938	0.947	0.947	0.902	0.952	0.910	0.951	0.949	0.937	0.950
0.2	0.2	5	40	0.799	0.951	0.947	0.801	0.958	0.859	0.971	0.957	0.937	0.955
0.2	0.2	5	60	0.899	0.940	0.940	0.874	0.946	0.900	0.957	0.953	0.938	0.954
0.2	0.2	5	80	0.937	0.944	0.946	0.905	0.952	0.910	0.948	0.947	0.937	0.948
0.2	0.1	25	40	0.532	0.965	0.953	0.582	0.979	0.661	0.976	0.960	0.954	0.955
0.2	0.1	25	60	0.684	0.958	0.950	0.726	0.965	0.774	0.959	0.952	0.951	0.948
0.2	0.1	25	80	0.797	0.953	0.953	0.821	0.960	0.856	0.958	0.953	0.949	0.951
0.2	0.1	15	40	0.542	0.959	0.950	0.595	0.974	0.669	0.972	0.954	0.951	0.950
0.2	0.1	15	60	0.691	0.959	0.954	0.730	0.966	0.785	0.963	0.954	0.952	0.949
0.2	0.1	15	80	0.795	0.951	0.950	0.812	0.953	0.853	0.959	0.953	0.947	0.947
0.2	0.1	5	40	0.547	0.960	0.950	0.601	0.977	0.659	0.975	0.960	0.954	0.957
0.2	0.1	5	60	0.696	0.958	0.951	0.738	0.965	0.789	0.962	0.954	0.952	0.950
0.2	0.1	5	80	0.799	0.949	0.946	0.827	0.955	0.864	0.957	0.952	0.948	0.949

Table 3. Results of the modified Poisson and least-squares regression analyses for the epilepsy clinical study ($N = 44$).

	Age at follow up	Gender (male vs. female)	Mood disorder symptoms	Graduating from school for special needs education
Modified Poisson regression				
Quasi-ML estimate	1.030	1.972	2.216	0.548
Ordinary Wald 95%C.I.	(0.983, 1.079)	(0.675, 5.762)	(0.492, 9.972)	(0.127, 2.377)
CQLR 95%C.I.	(0.975, 1.102)	(0.721, 8.776)	(0.580, 8.801)	(0.136, 2.407)
QLR (bootstrap) 95%C.I.	(0.975, 1.103)	(0.752, 7.985)	(0.594, 8.801)	(0.131, 2.387)
MAQLR 95%C.I.	(0.972, 1.094)	(0.761, 7.089)	(0.700, 5.743)	(0.200, 2.588)
MAQLR (bootstrap) 95%C.I.	(0.968, 1.119)	(0.675, ∞)	(0.553, 9.884)	(0.030, 3.297)
Modified least-squares regression				
Quasi-ML estimate	0.011	0.184	0.538	−0.189
Ordinary Wald 95%C.I.	(−0.002, 0.023)	(−0.066, 0.435)	(0.149, 0.928)	(−0.586, 0.207)
CQLR 95%C.I.	(−0.003, 0.032)	(−0.035, 0.528)	(0.025, 1.000)	(−1.000, 0.155)
QLR (bootstrap) 95%C.I.	(−0.003, 0.032)	(−0.031, 0.529)	(0.022, 1.000)	(−0.727, 0.152)
MAQLR 95%C.I.	(−0.003, 0.023)	(−0.029, 0.369)	(0.057, 0.784)	(−0.552, 0.133)
MAQLR (bootstrap) 95%C.I.	(−0.003, 0.032)	(−0.029, 0.613)	(0.027, 1.000)	(−0.635, 0.158)