Accurate inference methods based on the estimating equation theory for the modified Poisson and least-squares regressions

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

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

Objectives: In clinical and epidemiological studies, the modified Poisson and least-squares regression analyses for binary outcomes have been used as standard multivariate analysis methods to provide risk ratio and risk difference estimates. However, their ordinary Wald-type confidence intervals can suffer from finite-sample biases in the robust variance estimators, and the coverage probabilities of true effect measures are substantially below the nominal level (usually 95%). To address this issue, new accurate inference methods are needed.

Methods: We propose two accurate inference methods based on the estimating equation theory for these regression models. A remarkable advantage of these regression models is that the correct models to be estimated are known, that is, conventional binomial regression models with log and identity links. Using this modeling information, we first derive the quasi-score statistics, whose robust variances are estimated using the correct model information, and then propose a confidence interval based on the regression coefficient test using χ^2-approximation. To further improve the large sample approximation, we propose adapting a parametric bootstrap method to estimate the sample distribution of the quasi-score statistics using the correct model information. In addition, we developed an R package, `rqlm` (https://doi.org/10.32614/CRAN.package.rqlm), that can implement the new methods via simple commands.

Results: In extensive simulation studies, the coverage probabilities of the two new methods clearly outperformed the ordinary Wald-type confidence interval when the regression function assumptions were correctly specified, especially in small and moderate sample settings. We also illustrated the proposed methods by applying them to an epidemiological study of epilepsy. The proposed methods provided wider confidence intervals, reflecting statistical uncertainty.

Conclusions: The current standard Wald-type confidence intervals may provide misleading evidence. Erroneous evidence can potentially influence clinical practice, public health, and policymaking. These possibly inaccurate results should be circumvented using effective statistical methods. These new inference methods would provide more accurate evidence for future medical studies.

Key words: epidemiologic regression, generalized linear model, quasi-likelihood, estimating equation, bootstrap.

Introduction

Logistic regression has been a standard multivariate analysis method for analyzing binary outcome data in clinical and epidemiological studies. However, the odds ratio is difficult to interpret as an effect measure. While the causal odds ratio can be formally defined using a counterfactual framework, it can only be interpreted as an approximation of the risk ratio when the frequency of events is small [1, 2]. Thus, the use of risk ratios and risk differences is recommended as an alternative in various guidelines; for example, the CONSORT statements recommend reporting relative and absolute measures of effect when reporting the results of clinical trials [3, 4].

Owing to the substantial limitations of logistic regression, other binomial regression models have conventionally been considered using the log or identity link functions to provide risk ratio and risk difference estimators [5]. However, the values of these binomial regression models are not limited within the range [0, 1], and the maximum likelihood (ML) estimates often cannot be defined in practice [6, 7]. To address these issues, Zou [8] and Cheung [9] proposed modified Poisson and least-squares (Gaussian) regression analyses, which provide consistent risk ratio and risk difference estimators without computational difficulties. Their ideas were to formally fit the Poisson and least -squares regression models to the binary outcome data and calculate the regression coefficient estimates by the framework of the generalized linear model (GLM) [10]. The resultant estimators then provide consistent estimates of risk ratios and risk differences based on the estimating equation theory of the GLM [11, 12], even if the distributional assumptions are misspecified. In addition, the variance estimators should be changed to the sandwich variance estimators [13].

One relevant issue in logistic regression is the serious bias of the regression coefficient estimator for small samples, and various correction methods have been

discussed [14, 15]. Recently, Uno et al. [16] have shown that the same bias can occur in the modified Poisson regression analysis, although this phenomenon does not occur in the modified least-squares regression. More importantly, they showed that the robust variance estimators for both regression analysis methods can be biased and that the resultant Wald-type confidence intervals can seriously underestimate the actual statistical errors in small or moderate sample settings. These properties can lead to misleading evidence in clinical and epidemiological studies; therefore, accurate alternative statistical inference methods are needed.

In this study, we propose new confidence intervals for the modified Poisson and least-squares regression analyses based on the estimating equation theory, especially for accurate inferences in small or moderate sample settings. A remarkable advantage of these regression models is that we know the correct models to be estimated, that is, conventional binomial regression models with log and identity links. Using this modeling information, we first derive the quasi-score statistics for these regression models, whose robust variances are estimated using the correct model information. Quasi-score-based inferences have been discussed for various pseudo-likelihood inferences (e.g., Mantel–Haenszel methods [17, 18]) and are known to have favorable properties compared with naïve Wald-type inferences. We subsequently propose a confidence interval based on the quasi-score test using χ^2 -approximation. To further improve small sample approximations, we propose adapting a parametric bootstrap method to estimate the sample distribution of the quasi-score statistics using the correct model information. Through extensive simulation studies, we showed that the coverage probabilities of the two new methods clearly outperform the ordinary Wald-type confidence interval. We also illustrate the proposed methods through their application in an epidemiological study of epilepsy. We developed an R package, `rqlm`

(https://doi.org/10.32614/CRAN.package.rqlm), that can implement the new methods via simple commands.

Modified Poisson and least-squares regressions

We consider a cohort study consisting of n participants with binary outcomes $Y_1, \dots, Y_n$ ($= 1$: event occurred, $= 0$: event did not occur) and covariates $\boldsymbol{x}_i = \left(x_{i1}, x_{i2}, \dots, x_{ip}\right)^T$ for the ith subject ($i = 1, \dots, n$). Conventionally, binomial regressions with log link and identity link functions have been considered for multivariate analyses of risk ratios and risk differences; however, they involve serious theoretical difficulties in defining the ML estimates because the values of the regression functions do not fall within [0, 1] [6, 7]. Zou [8] and Cheung [9] proposed the modified Poisson and least-squares regressions as effective methods for multivariate analyses. Their ideas involved formally fitting the Poisson and least-squares regression models 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 $\boldsymbol{\beta} = \left(\beta_0, \beta_1, \dots, \beta_p\right)^T$ become consistent estimators of the log-transformed risk ratios and risk differences on the target population [16]. The principle of these estimating methods is based on the estimating equation theory of the GLM [12]; that is, the estimating functions are unbiased even if the distribution forms are misspecified as long as the functional forms of the regression functions are correctly specified. In particular, for the modified least-squares regression, the quasi-ML estimator becomes a linear unbiased estimator [16]; $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator for the regression coefficients of the binomial regression function [16]. The standard errors of $\widehat{\boldsymbol{\beta}}$ of both models are consistently estimated by the sandwich variance estimator [13].

Confidence intervals for risk ratios and risk differences

Confidence intervals based on the quasi-score statistics

The modified Poisson and least-squares regressions are effective methods for multivariate analyses of risk ratios and risk differences; however, their ordinary Wald-type confidence intervals can seriously underestimate statistical errors in small or moderate sample settings [16]. To address these issues, we first derive quasi-score tests for the regression coefficients. The two models are formulated as specific cases of the GLM, and the quasi-likelihood estimating functions are expressed as

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

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 $\boldsymbol{D}_i = \partial \mu_i / \partial \boldsymbol{\beta}$; moreover, $V_i = v(\mu_i)$ is the variance function of the outcome variable ($= \mu_i$ for the Poisson model and $= 1$ for the Gaussian model) ($i = 1, \dots, n$). If the variance functions are correctly specified, the covariance matrices of $U(\boldsymbol{\beta})$ become the Fisher information matrices $I(\boldsymbol{\beta}) = -E[\partial U(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}]$. However, the variance functions are misspecified for these cases; therefore, they become the robust covariance matrices, $J(\boldsymbol{\beta}) = E[U(\boldsymbol{\beta}) U^T(\boldsymbol{\beta})]$. In addition, a special feature of these inferences is that the correct models are known to us and expectations can be substituted into the binomial regression models. Then, the concrete forms of the models are expressed as

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{x}_i (Y_i - \mu_i) = \boldsymbol{X}(\boldsymbol{Y} - \boldsymbol{\mu})$$

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

where $\boldsymbol{Y} = (Y_1, \dots, Y_n)^T, \boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T, \boldsymbol{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$, and $\boldsymbol{W} = \mathrm{diag}\{\mu_1(1 -$

$\mu_1), \dots, \mu_n(1-\mu_n)\}$. Although the definitions of the mean function μ_i differ between the two models, the function forms are the same. Note that $\mu_1, \dots, \mu_n$ should be truncated on [0, 1] on $\boldsymbol{W}$ because the individual variance functions substantially estimate the variances of the binomial variables and should not be negative values. However, those in $U(\boldsymbol{\beta})$ should not be truncated; if they are truncated, the quasi-score functions are biased and unrealistically singular results can be obtained. Furthermore, we note that $E[U(\boldsymbol{\beta})] = \mathbf{0}$ and $V[U(\boldsymbol{\beta})] = J(\boldsymbol{\beta})$ without large sample approximations. The quasi-score test statistics for the joint null hypotheses H_0: $\boldsymbol{\beta} = \boldsymbol{\beta}_{\text{null}}$ are constructed using the exact means and covariance matrices. However, these null hypotheses are usually outside the scope of interest in practice.

We consider tests for composite null hypotheses H_0: $\beta_1 = \beta_{1,\text{null}}$ that correspond to the hypothesis tests for individual risk ratios and risk differences; without loss of generality, we consider the tests of regression coefficients of first explanatory variables and denote $\boldsymbol{\beta}_c = \left(\beta_2, \beta_3, \dots, \beta_p\right)^T$. The quasi-score statistics are then constructed as follows:

$$T(\beta_{1,\text{null}}) = U^T(\widetilde{\boldsymbol{\beta}})J^{-1}(\widetilde{\boldsymbol{\beta}})U(\widetilde{\boldsymbol{\beta}})$$

where $\widetilde{\boldsymbol{\beta}} = \left(\tilde{\beta}_0, \beta_{1,\text{null}}, \widetilde{\boldsymbol{\beta}}_c\right)^T$; $\tilde{\beta}_0$ and $\widetilde{\boldsymbol{\beta}}_c = \left(\tilde{\beta}_2, \tilde{\beta}_3, \dots, \tilde{\beta}_p\right)^T$ are the constrained quasi-ML estimates of $\{\beta_0, \boldsymbol{\beta}_c\}$ under H_0. The constrained quasi-ML estimates can be calculated via the same modified Poisson and least-squares regression analyses when x_{1i} is dropped from the explanatory variables and offsets $\beta_{1,\text{null}}x_{1i}$ are added ($i = 1, \dots, n$). Under the null hypotheses, the quasi-score test statistics follow the χ^2-distribution with approximately one degree of freedom [19]. Moreover, using these quasi-score tests, we can construct the $100 \times (1-\alpha)\%$ confidence intervals of β_1 using the sets of null values that satisfy

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

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 be calculated using adequate numerical methods (e.g., the bisectional methods [20]).

As shown in the simulation studies, confidence intervals based on the quasi-score statistics generally have favorable properties compared with the ordinary Wald-type confidence intervals obtained using the standard sandwich variance estimators. However, confidence intervals when using the large sample χ^2 approximations still have limitations with respect to achieving sufficient coverage performance in small sample settings [21]. In addition, more accurate approximations of the sample distributions are needed for valid inferences.

Bootstrap confidence intervals based on the quasi-score statistics

To improve the accuracy of approximations of the sample distributions of the quasi-score statistics, we propose using the bootstrap method. Again, we focus on the advantage of these regression methods—that the correct distributional assumptions for the target population are known (binomial regression models). Therefore, we propose performing parametric bootstrap resampling from the "correct" binomial regression models by substituting the regression coefficients $\boldsymbol{\beta}$ for the null value and constrained quasi-ML estimates. The bootstrap algorithm for the tests of H_0: $\beta_1 = \beta_{1,\text{null}}$ is as follows.

Algorithm (bootstrap tests for the quasi-score statistics)

1. For the modified Poisson and least-squares regression models, compute the constrained quasi-ML estimates $\tilde{\beta}_0$ and $\tilde{\boldsymbol{\beta}}_c$ under H_0: $\beta_1 = \beta_{1,\text{null}}$.

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

$$\Pr(Y_i = 1|\boldsymbol{x}_i) = \exp(\tilde{\beta}_0 + \beta_{1,\text{null}}x_{i1} + \tilde{\beta}_2 x_{i2} + \cdots + \tilde{\beta}_p x_{ip})$$

$$\Pr(Y_i = 1|\boldsymbol{x}_i) = \tilde{\beta}_0 + \beta_{1,\text{null}}x_{i1} + \tilde{\beta}_2 x_{i2} + \cdots + \tilde{\beta}_p x_{ip}$$

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

3. Compute the quasi-score statistic $T^{(b)}(\beta_{1,\text{null}})$ for the b th bootstrap sample $Y_1^{(b)}, \ldots, Y_n^{(b)}$ $(b = 1,2, \ldots, B)$.

4. Calculate the empirical distribution function of $T^{(1)}(\beta_{1,\text{null}}), \ldots, T^{(B)}(\beta_{1,\text{null}})$ — specifically, $\bar{F}_{T(\beta_{1,\text{null}})}(t)$—which is the bootstrap estimate of the sample distribution of $T(\beta_{1,\text{null}})$.

5. Implement the hypothesis test for $T(\beta_{1,\text{null}})$ using $\bar{F}_{T(\beta_{1,\text{null}})}(t)$ for the reference distribution.

Because accurate tail area estimation of the null distribution requires a large number of replications [22], the number of bootstrap resamplings B should be sufficiently large (usually at least 1000).

The corresponding $100 \times (1 - \alpha)\%$ confidence intervals of β_1 can be constructed using the sets of $\beta_{1,\text{null}}$ that satisfy

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

The confidence limits can also be calculated using adequate numerical methods (e.g., the bisectional methods [20]). The parametric bootstrap approach effectively uses the distributional information of the correct models; thus, the approximation is expected to improve compared with the naïve asymptotic normal approximation. The actual performances are demonstrated in simulation studies. Note that the validity of the

proposed confidence intervals requires correct specifications of the regression functions, which can be violated if the pivotal assumptions are incorrect. However, the validity of other standard confidence intervals (e.g., Wald-type and nonparametric bootstrap confidence intervals) is also violated under these conditions.

Software

We developed an R package, `rqlm` (https://cran.r-project.org/web/packages/rqlm), to perform all of the proposed methods via simple commands. An example of the R code used to implement these methods is provided in the Supplementary Materials.

Simulations

To illustrate the operating characteristics of the proposed methods, we conducted simulation studies. For data generation, we considered the binomial regression models with log and identity link functions, and parameter settings were selected to mimic the epidemiological study of epilepsy described in the next section. Four explanatory variables were considered: x_{i1} was the main treatment/exposure variable that followed a Bernoulli distribution with probability 0.20 or 0.10; x_{i2} was a confounding variable that followed a Bernoulli distribution with probability 0.773 and had a correlation with x_{i1} through measurement of the odds ratio [OR] $\Pr(x_{i1} = 1, x_{i2} = 1)\Pr(x_{i1} = 0, x_{i2} = 0)/\Pr(x_{i1} = 0, x_{i2} = 1)\Pr(x_{i1} = 1, x_{i2} = 1)$ = 25, 15, and 5; x_{i3} followed a Bernoulli distribution with probability 0.455; and x_{i4} followed $N(29.0, 7.37)$. The outcome variable Y_i was then 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 for the overall event rate of the cohort; the event rate was varied as 0.40 and 0.20. 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 former model and $= (0.116, -0.041, 0.0037, 0.023)$ for the latter model. We considered the sample size $n = 20, 30, \ldots, 100$; small and moderate sample settings in which the robust variance estimator may not perform well. If separation or quasi-separation situations occur, the corresponding cases are excluded from the experiments because the quasi-ML estimates might not be obtained under these settings. We performed 2,000 simulations for the 108 scenarios for both the risk ratio and risk difference regression models.

For comparisons, we analyzed the individual dataset using the ordinary modified Poisson and least-squares regressions and Wald-type confidence intervals obtained via the standard sandwich variance estimator, HC3-type bias-adjusted sandwich variance estimator [23], Firth-type correction method using the bias-adjusted sandwich variance estimator [16], nonparametric bootstrap confidence intervals [22], and Fisher's *z*-transformation proposed by Zou and Donner [24] (only for the modified least-squares regression). We then applied the two proposed confidence intervals based on the quasi-score statistics and bootstrap approach. For the bootstrap methods, we performed 2,000 bootstrap resamplings to calculate the bootstrap distributions. We assessed the coverage probabilities of the 95% confidence intervals for β_1 of the three methods. The R codes used to implement the simulations are provided in the Supplementary Materials.

The results of the simulation studies are presented in Figures 1 and 2 for the modified Poisson regression and in Figures 3 and 4 for the modified least-squares regression. The empirical coverage rates of the 95% confidence intervals for the 2,000 simulations were plotted. For the modified Poisson regression, the four conventional confidence intervals suffered from undercoverage in small and moderate sample settings. These results were

due to the biases of both the regression coefficient estimates and robust standard error estimates. The improved methods based on higher-order approximations (HC3 and Firth-type correction) still suffered from invalidity in these settings. The degree of undercoverage was serious if the event rate or exposure rate became too small and the correlations of x_{i1} and x_{i2} became too large. In addition, the two proposed confidence intervals retained coverage probabilities around the nominal level (95%) for most scenarios. The quasi-score-based confidence intervals could reflect undercoverage bias in some small sample settings (generally, $n \leq 50$). However, the coverage properties were consistently favorable compared with those of the conventional confidence interval. In addition, as expected, the proposed bootstrap confidence intervals exhibited better coverage probabilities. Even for small sample settings ($n = 20$), the coverage probabilities consistently had values around the nominal level (95%).

For the modified least-squares regression, the overall simulation results were similar to those of the modified Poisson regression simulations. The conventional confidence intervals were generally prone to undercoverage in small and moderate samples. The quasi-ML estimates corresponded to a linear unbiased estimator [16]; thus, there were no biases in the point estimator. However, robust variance estimators were seriously biased in small and moderate sample settings. Moreover, the degree of undercoverage became serious if the event rate or exposure rate became excessively small and the correlations of x_{i1} and x_{i2} became large. In addition, the proposed confidence intervals exhibited favorable coverage properties. The quasi-score-based confidence interval was prone to undercoverage in some settings; however, the coverage properties were much better than those of the conventional methods. Additionally, the proposed bootstrap confidence interval consistently improved the coverage properties. Even in small sample settings ($n = 20$), the coverage probabilities consistently had values close to the nominal level

(95%). These simulation results indicate that the proposed bootstrap confidence interval provided accurate interval estimates and that validity was retained even in small and sparse data settings.

Applications

To illustrate the usefulness of the proposed methods, we applied them to an epidemiological study of epilepsy conducted by Arai et al. [25], which was a retrospective cohort study that evaluated the factors associated with the employment status of patients with a history of childhood-onset drug-resistant epilepsy (N = 56). We analyzed cohort data using the modified Poisson and least-squares regressions. The outcome was employment status (1 = non-employment, 0 = employment; the number of events was 14), and four explanatory variables were included: age at follow-up (`age`; continuous), gender (`gender`; male or female), mood disorder symptoms (`symptoms`; yes or no), and graduating from a school for special needs education (`education`; yes or no). The last two variables were highly correlated and were strongly associated with the outcome through univariate analyses [25].

Table 1 presents the results of the study. In many cases, the confidence intervals based on the quasi-score statistics and bootstrap approach were asymmetric around the quasi-ML estimates. In some cases, the locations of the confidence intervals differed substantially from those of the Wald-type confidence intervals. This phenomenon is known to occur in efficient score-based confidence intervals [21]. For the modified Poisson regression, only the 95% Wald-type confidence interval for `education` did not cover the null value (= 1). The two proposed confidence intervals also did not cover 1; however, the bootstrap confidence interval was relatively narrow compared with the Wald-type confidence interval. For `symptoms`, the confidence interval based on the

quasi-score statistic was narrower than the Wald-type confidence interval and did not cover 1; however, the bootstrap confidence interval was much wider and did not cover 1. These results might be influenced by the strong correlation between these two covariates. In addition, for the modified least-squares regression, the Wald-type confidence interval of `symptoms` was substantially widened using the bootstrap-based approach. Additionally, the bootstrap confidence interval for `education` was wider than the Wald-type confidence interval. Considering the operating characteristics shown by the simulations, the conclusions obtained from the ordinary Wald-type confidence intervals might be misleading, and improved methods would likely provide more precise evidence.

Discussion

The modified Poisson and least-squares regressions have been widely used in recent clinical and epidemiological studies because they provide interpretable effect measure estimates without computational difficulties. Considering the difficulty in interpreting the odds ratios, these methods will be increasingly adopted in future studies as effective alternatives to the conventional logistic regression. This study demonstrates how ordinary Wald-type confidence intervals in these settings are prone to serious undercoverage with small or moderate sample sizes. As clearly shown in the real data example, the ordinary confidence intervals might provide misleading evidence, which can potentially influence clinical practice, public health, and policymaking. These possibly inaccurate results should be circumvented using effective statistical methods. The new accurate confidence intervals and numerical evidence provided in this study will be useful for future medical studies.

In the simulation-based evidence, the ordinary Wald-type and other conventional confidence intervals showed serious undercoverage performance. For the modified

Poisson regression, a relevant reason for the invalid property is the small sample bias; although bias of the ML estimator for the GLM is well known [26], a similar bias can occur for the quasi-ML estimator for misspecified models. Effective solutions to address this bias are the Firth-type bias correction [16] and higher-order bias correction methods [26]. In addition, for the modified least-squares regression, the resultant estimator is unbiased because it corresponds to a linear unbiased estimator, and no corrections are required if this criterion is considered [16]. However, the undercoverage properties shown in the simulation studies are severe, and bias of the ordinary robust variance estimator should be adequately addressed in practice. The proposed methods are expected to provide an effective solution to this problem.

A fundamental assumption to ensure consistency of the quasi-ML estimators is that the regression models are correctly specified, similar to other regression models involving ordinary logistic regression analysis. In addition, the current available inference methods employed in the simulation studies assume that the regression models are correctly specified. This is a practically unverifiable assumption but should be carefully considered in practice. In particular, misspecifications of the regression models can violate the validity of the proposed confidence intervals, and the validity of the variance estimators is violated if the regression models are misspecified. Similar problems can occur for other current standard confidence intervals (e.g., Wald-type and nonparametric bootstrap confidence intervals), and further developments of alternative robust methods to address the model misspecification problem are relevant issues in future studies. An effective approach to resolve this issue is to adopt nonlinear regression models (e.g., spline models), especially in cases where continuous covariates are modelled [27]. In addition, other regression analysis methods have been proposed to estimate interpretable effect measures [28-31], and similar discussions are beneficial for these alternative effective approaches.

Developments of new methods for these alternative regression methods would also be relevant future issues.

Another future issue is adapting improved robust variance estimators for the Wald-type confidence interval [32]. Although the accuracy of the existing improved robust variance estimators is better than that of the ordinary sandwich variance estimator, deterministic conclusions cannot be provided for their relative performances [32] because all methods are based on approximations (e.g., higher-order approximations). Although simulation-based numerical evidence can provide good case-by-case comparative performance, they would not be generic properties. However, the bootstrap-based approach proposed in the present study is based on the quasi-score statistic, which effectively uses information from the null hypothesis and adapts the flexible sample distribution estimate. Therefore, we believe that it is one of the most effective methods among the competing methods. The proposed methods can be used as accurate and effective alternatives to the ordinary Wald-type inference methods in clinical and epidemiological studies.

Acknowledgements

The authors would like to thank Dr. Y. Arai (Tottori University) for permission to use the valuable data and K. Nakazono (The Institute of Statistical Mathematics) for his helpful comments on the earlier draft.

Competing interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

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).

Data availability

R package for implementing the proposed methods is available at CRAN (https://cran.r-project.org/web/packages/rqlm).

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. Euro J Epidemiol. 1995;11:365-371.
3. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 statement: updated guideline for reporting randomized trials. Nat Med. 2025. DOI: 10.1038/s41591-025-03635-5.
4. Thompson J, Watson SI, Middleton L, et al. Estimating relative risks and risk differences in randomised controlled trials: a systematic review of current practice. Trials. 2025;26:1.
5. Rothman KJ, Greenland G, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
6. McNutt LA, Wu C, Xue X, et al. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am J Epidemiol. 2003;157:940-3.
7. 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.

8. Zou GY. A modified Poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004;159:702-6.
9. Cheung YB. A modified least-squares regression approach to the estimation of risk difference. Am J Epidemiol. 2007;166:1337-44.
10. Nelder JA, Wedderburn RWM. Generalized linear models. J Royal Stat Soc A. 1972;135:370-384.
11. Godambe VP, Heyde CC. Quasi-likelihood and optimal estimation. Int Stat Rev. 1987;55:231-244.
12. Wedderburn RWM. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika. 1974;61:439-447.
13. White H. Maximum likelihood estimation of misspecified models. Econometrica. 1982;50:1-25.
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. 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.
17. Sato T. Confidence limits for the common odds ratio based on the asymptotic distribution of the Mantel-Haenszel estimator. Biometrics. 1990;46:71-80.
18. Noma H, Nagashima K. A note on the Mantel-Haenszel estimators when the common effect assumptions are violated. Epidemiol Methods. 2016;5:19-35.
19. Cox DR, Hinkley DV. Theoretical Statistics. London: Chapman and Hall; 1974.

20. Burden RL, Faires JD. Numerical Analysis. 9th ed ed. Boston: Cengage Learning; 2011.
21. Noma H. Confidence intervals for a random-effects meta-analysis based on Bartlett-type corrections. Stat Med. 2011;30:3304-3312.
22. Efron B, Tibshirani R. An Introduction to the Bootstrap. New York: CRC Press; 1994.
23. White H. A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity. Econometrica. 1980;48:817-838.
24. Zou G, Donner A. A simple alternative confidence interval for the difference between two proportions. Control Clin Trials. 2004;25:3-12.
25. 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.
26. Cordeiro GM, McCullagh P. Bias correction in generalized linear models. J Royal Stat Soc B. 1991;53:629-643.
27. Noma H, Kitano T. Modelling nonlinear effects in risk ratio and risk difference using the Poisson and Gaussian additive regression models. Stats. 2024;7:1473-1482.
28. Diaz-Quijano FA. A simple method for estimating relative risk using logistic regression. BMC Med Res Methodol. 2012;12:14.
29. Dwivedi AK, Mallawaarachchi I, Lee S, et al. Methods for estimating relative risk in studies of common binary outcomes. J Appl Stat. 2014;41:484-500.
30. Richardson TS, Robins JM, Wang L. On modeling and estimation for the relative risk and risk difference. J Am Stat Assoc. 2017;112:1121-1130.
31. Talbot D, Mesidor M, Chiu Y, et al. An alternative perspective on the robust Poisson method for estimating risk or prevalence ratios. Epidemiology. 2023;34:1-7.
32. Gosho M, Ishii R, Noma H, et al. A comparison of bias-adjusted generalized

estimating equations for sparse binary data in small-sample longitudinal studies. Stat Med. 2023;42:2711-2727.

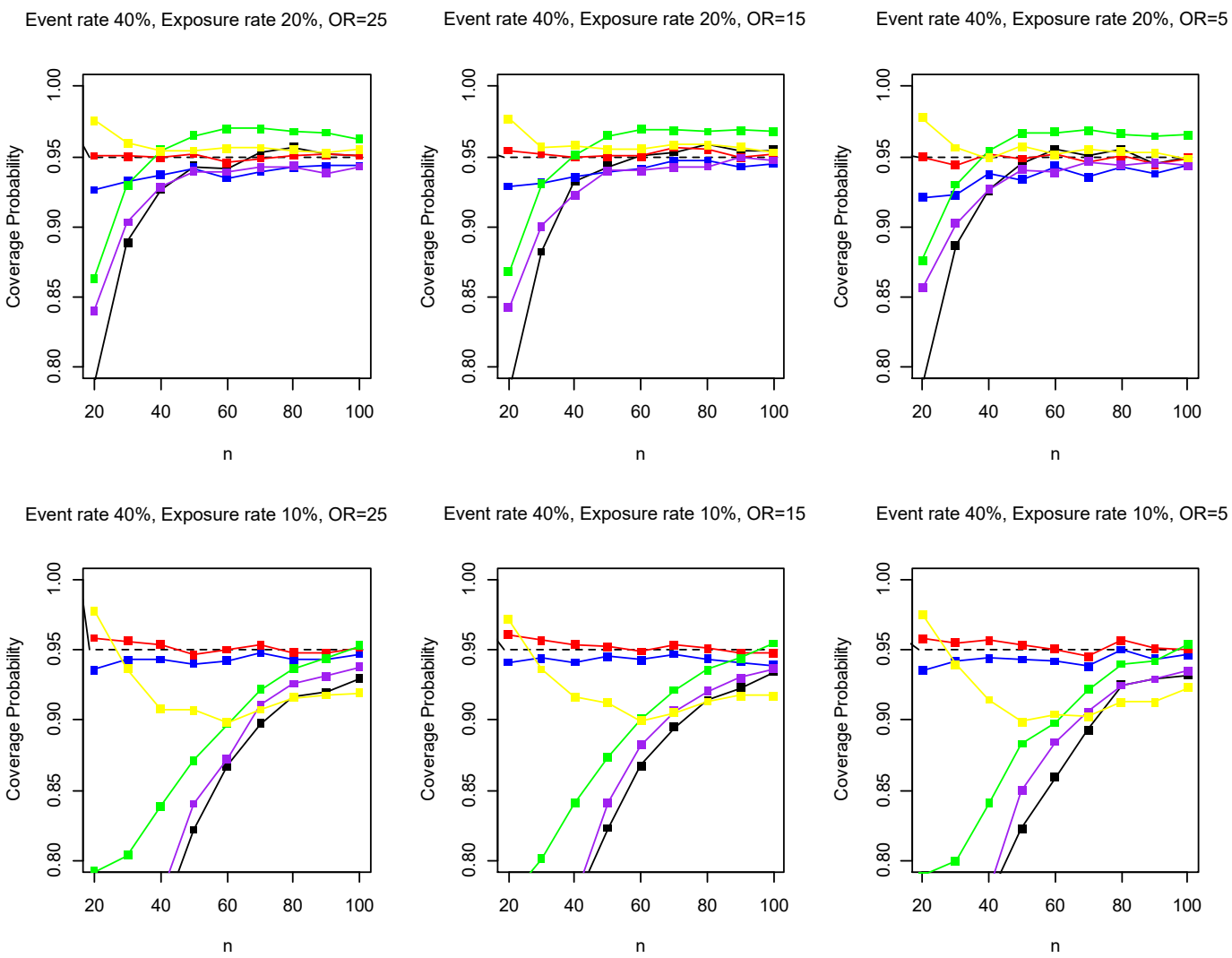

Figure 1. Results of simulations for the modified Poisson regression (**black**: ordinary Wald CI, **green**: Wald CI using HC3-estimator, **yellow**: Wald CI using Firth-type estimator, **purple**: nonparametric bootstrap CI, **blue**: quasi-score CI, **red**: bootstrap CI by the quasi-score statistic; CI: confidence interval).

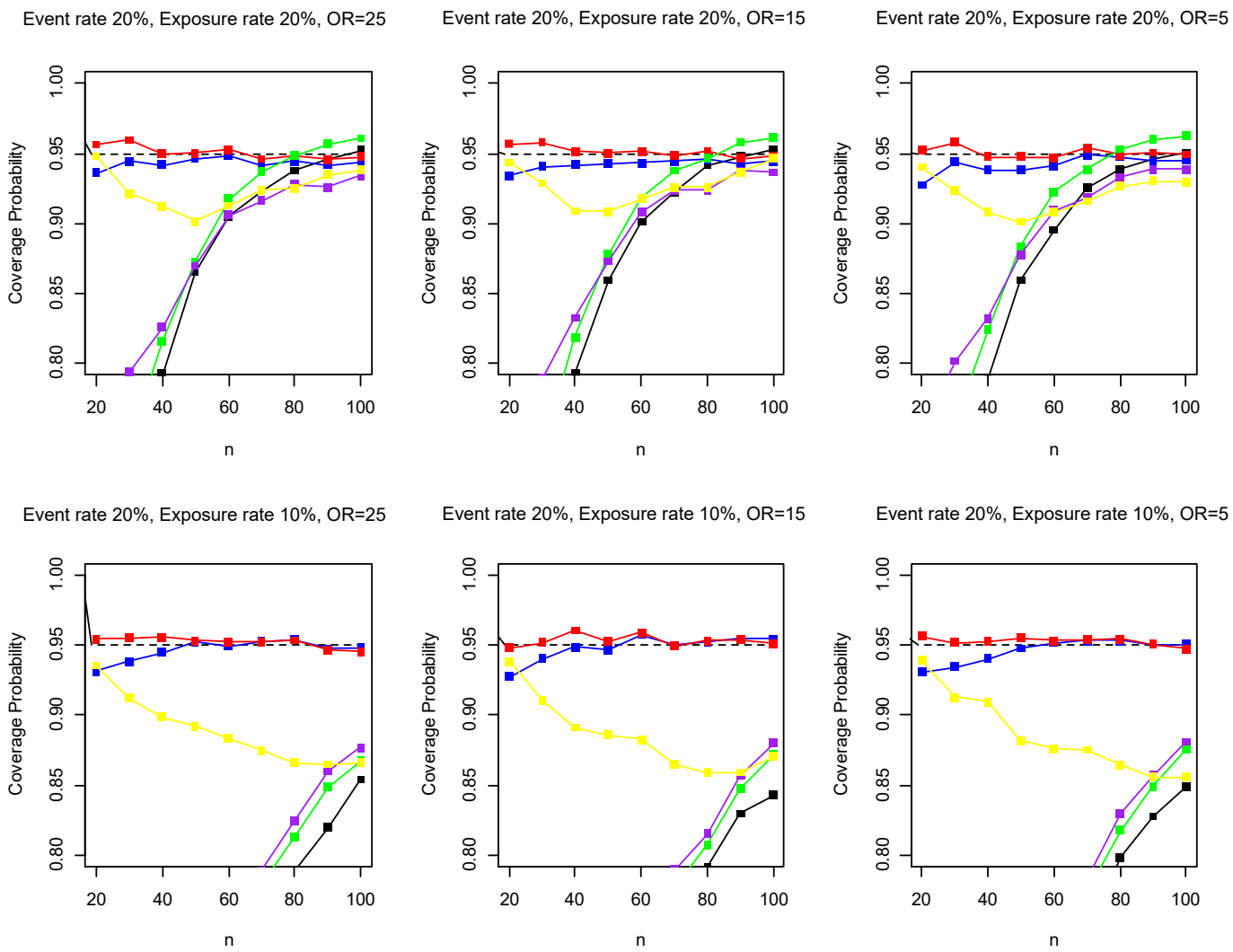

Figure 2. Results of simulations for the modified Poisson regression (**black**: ordinary Wald CI, **green**: Wald CI using HC3-estimator, **yellow**: Wald CI using Firth-type estimator, **purple**: nonparametric bootstrap CI, **blue**: quasi-score CI, **red**: bootstrap CI by the quasi-score statistic; CI: confidence interval).

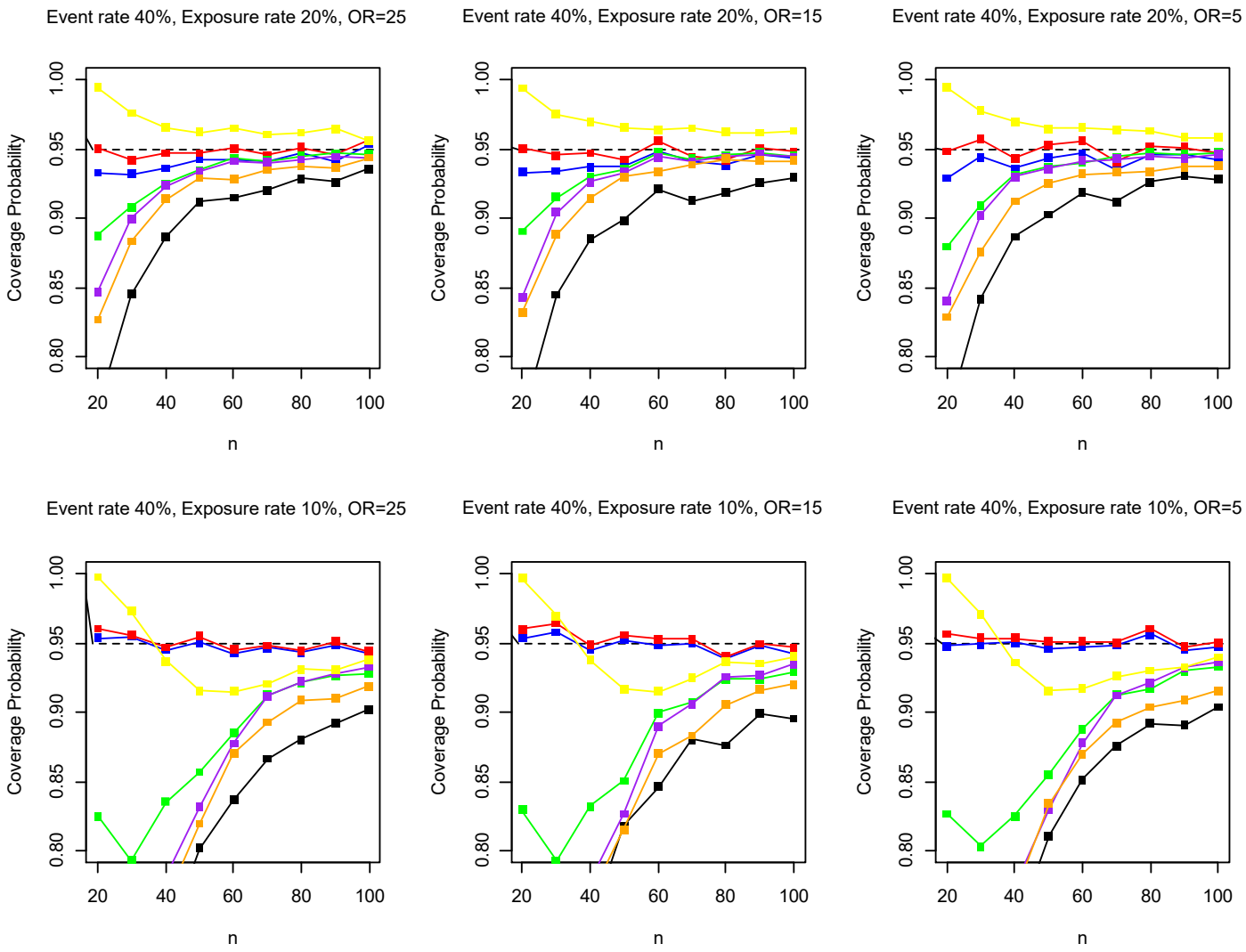

Figure 3. Results of simulations for the modified least-squares regression (**black**: ordinary Wald CI, **green**: Wald CI using HC3-estimator, **yellow**: Wald CI using Firth-type estimator, **purple**: nonparametric bootstrap CI, **orange**: Fisher's *z*-transformation, **blue**: quasi-score CI, **red**: bootstrap CI by the quasi-score statistic; CI: confidence interval).

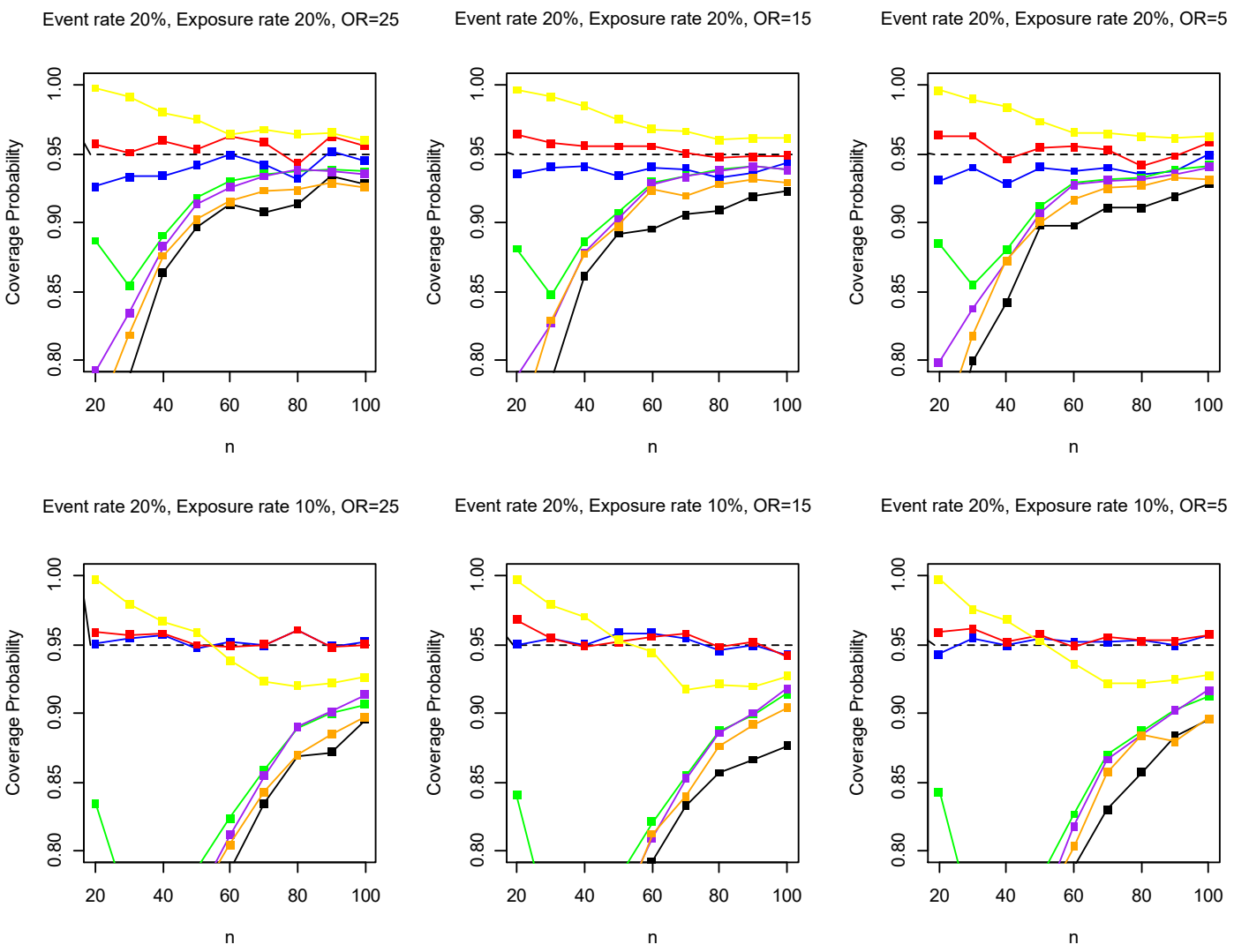

Figure 4. Results of simulations for the modified least-squares regression (**black**: ordinary Wald CI, **green**: Wald CI using HC3-estimator, **yellow**: Wald CI using Firth-type estimator, **purple**: nonparametric bootstrap CI, **orange**: Fisher's *z*-transformation, **blue**: quasi-score CI, **red**: bootstrap CI by the quasi-score statistic; CI: confidence interval).

Table 1. Results of the modified Poisson and least-squares regression analyses for the epilepsy epidemiological study ($N = 56$) †.

	Risk ratio estimation using the modified Poisson regression			Risk difference estimation using the modified least-squares regression		
	Quasi-ML estimate (Wald 95%CI)	95%CI by the quasi-score statistic	Bootstrap 95%CI by the quasi-score statistic	Quasi-ML estimate (Wald 95%CI)	95%CI by the quasi-score statistic	Bootstrap 95%CI by the quasi-score statistic
Age at follow-up	1.013 (0.982, 1.045)	(0.981, 1.050)	(0.978, 1.067)	0.006 (−0.003, 0.014)	(−0.003, 0.014)	(−0.003, 0.022)
Gender (male vs. female)	1.372 (0.638, 2.950)	(0.669, 3.030)	(0.614, 3.509)	0.097 (−0.089, 0.283)	(−0.063, 0.237)	(−0.071, 0.426)
Mood disorder symptoms (yes vs. no)	2.076 (0.781, 5.519)	(1.001, 3.230)	(0.882, 7.385)	0.498 (0.153, 0.842)	(0.103, 0.721)	(0.067, 1.000)
Graduating from a school for special needs education (yes vs. no)	0.240 (0.075, 0.765)	(0.100, 0.686)	(0.028, 0.721)	−0.422 (−0.758, −0.087)	(−0.688, −0.098)	(−0.781, −0.086)

† CI: confidence interval.