DOI: 10.2478/bile-2026-0012

Biometrical Letters
Vol. 0 (2026), No. 0, 1-19

Exact unbiased covariance estimation for modified least-squares regression with binary outcomes

Takahiro Kitano[1], Hisashi Noma[1,2]

[1] The Graduate Institute for Advanced Studies, The Graduate University for Advanced Studies, Tokyo, Japan
[2] Department of Interdisciplinary Statistical Mathematics, The Institute of Statistical Mathematics, Tokyo, Japan, e-mail: noma@ism.ac.jp

SUMMARY

Logistic regression has long been widely used in clinical and epidemiologic studies with binary outcomes. However, because odds ratios can be difficult to interpret, recent publications have increasingly recommended alternative regression models that yield more directly interpretable effect measures. Modified least-squares regression is one such approach, providing covariate-adjusted estimates of risk differences. Its point estimator is the ordinary least-squares estimator, whereas statistical inference is typically based on heteroskedasticity-robust sandwich covariance estimators. Although these covariance estimators are asymptotically valid, they may be biased in small to moderate samples, particularly in the presence of high-leverage observations. In this article, we show that, under a correctly specified linear mean model with fixed covariates, the covariance matrix of the modified least-squares estimator admits an exactly unbiased finite-sample estimator. The proposed construction uses the Hadamard square of the residual-maker matrix to recover unbiased estimates of the individual Bernoulli variances from the expected squared residuals. The resulting covariance estimator is algebraically simple and can be computed using only standard least-squares quantities. We also discuss its limitations, including the possibility of negative individual variance estimates, failure of the estimated covariance matrix to be positive semidefinite, and numerical instability when the Hadamard matrix is ill-conditioned. Exact unbiasedness of the covariance estimator does not imply exact finite-sample Wald inference.

Key words: binary outcome; finite-sample bias; heteroskedasticity; modified least-squares regression; risk difference; sandwich covariance

1. Introduction

Logistic regression is widely used for multivariable analyses of binary outcomes in clinical and epidemiological research. Nevertheless, its principal effect

measure, the odds ratio, is often less intuitive than measures expressed directly on the probability scale; the odds ratio generally approximates the risk ratio only when the outcome is rare (Altman, Deeks and Sackett, 1998; Nurminen, 1995). Consequently, reporting effect measures such as risk ratios and risk differences has increasingly been encouraged. For example, the CONSORT guidelines recommend presenting both relative and absolute measures when reporting treatment effects in clinical trials (Hopewell et al., 2025; Thompson et al., 2025).

One approach to estimating covariate-adjusted risk differences is binomial regression with an identity link (Rothman, Greenland and Lash, 2008). In practice, however, maximum likelihood estimation must enforce the admissible probability range of 0 to 1 and may fail because of convergence or boundary problems (McNutt et al., 2003; Wallenstein and Bodian, 1987). As a computationally stable alternative, Cheung (2007) proposed modified least-squares regression, in which an ordinary Gaussian linear model is applied directly to a binary outcome. The regression coefficients can be estimated using standard least-squares or generalized linear model procedures (Nelder and Wedderburn, 1972). Under an appropriately specified mean structure, the resulting estimator consistently estimates covariate-adjusted risk differences, even though the assumed Gaussian variance structure is incorrect (Godambe and Heyde, 1987; Wedderburn, 1974). Valid statistical inference therefore requires a heteroskedasticity-robust sandwich covariance estimator rather than the conventional model-based covariance estimator (Eicker, 1963, 1967; Huber, 1967; White, 1980).

The finite-sample behavior of heteroskedasticity-robust covariance estimators has long been recognized as problematic. Several corrections, including HC1, HC2, HC3 and leverage-adjusted variants, have been developed to reduce bias and improve test performance (Chesher and Austin, 1991; Chesher and Jewitt, 1987; Cribari-Neto, 2004; Cribari-Neto, Souza and Vasconcellos, 2007; Hinkley, 1977; Long and Ervin, 2000; MacKinnon and White, 1985). Related issues also arise in generalized estimating equations and estimating-equation-based methods, where small-sample corrections to sandwich estimators

have been extensively studied (Fay and Graubard, 2001; Gosho, Sato and Takeuchi, 2014; Mancl and DeRouen, 2001; Morel, Bokossa and Neerchal, 2003; Pan and Wall, 2002; Wang and Long, 2011). Recent studies on modified Poisson and modified least-squares regression have likewise emphasized that ordinary robust standard errors and Wald-type intervals can be inaccurate in small or sparse samples (Noma and Gosho, 2025; Noma, Sunada and Gosho, 2025; Uno, Noma and Gosho, 2024).

This article revisits modified least-squares regression from a simple fixed-design least-squares perspective. The key observation is that the squared residuals are biased estimates of the individual Bernoulli variances because the residual-maker matrix mixes the error components. By inverting this mixing operation, one obtains a finite-sample unbiased estimator of the individual variances, and hence of the covariance matrix of $\widehat{\boldsymbol{\beta}}$. The estimator is a direct adaptation of the Hartley–Rao–Kiefer/Rao type unbiased variance construction in heteroskedastic linear models (Bera, Suprayitno and Premaratne, 2002; Dobriban et al., 2018; Hartley, Rao and Kiefer, 1969; Rao, 1970). The Hadamard construction itself is therefore not new; the contribution of this note is its explicit formulation, implementation, and finite-sample assessment for modified least-squares risk-difference regression with binary outcomes. Although the Bernoulli variance $v_i = p_i(1 - p_i)$ is determined by the mean, the Hadamard approach estimates the observation-specific variance components from residuals rather than relying on a plug-in mean–variance model, thereby retaining a sandwich-style heteroskedasticity correction at the cost of additional finite-sample variability. Here, "exact" refers to finite-sample unbiasedness of covariance estimation, not to exact finite-sample inference.

2. Modified least-squares regression

Suppose that $Y_1, \ldots, Y_n$ are binary variables given a fixed design matrix $\boldsymbol{X}$, with full column rank p. The argument below requires only that the regression errors are conditionally uncorrelated given $\boldsymbol{X}$; conditional independence is a convenient

sufficient condition in the binary-outcome setting. Let $\boldsymbol{x}_i^\top$ denote the i-th row of $\boldsymbol{X}$. Assume that the linear probability model is correctly specified:

$$\mathrm{E}(Y_i \mid \boldsymbol{X}) = p_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}_0,$$

where $0 \leq p_i \leq 1$. The parameter $\boldsymbol{\beta}_0$ is estimated by ordinary least squares, equivalently by fitting a Gaussian generalized linear model with an identity link (Nelder and Wedderburn, 1972).

Define $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ and

$$\mathrm{E}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \boldsymbol{0},$$

$$\mathrm{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \mathrm{diag}(v_1, \ldots, v_n), v_i = p_i(1 - p_i).$$

The modified least-squares estimator (Cheung, 2007) of the risk differences is

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}.$$

Because $\mathrm{E}(\mathbf{Y} \mid \mathbf{X}) = \boldsymbol{X}\boldsymbol{\beta}_0$, $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}_0$ in finite samples (Noma and Gosho, 2025; Noma et al., 2025; Uno et al., 2024).

Let $\boldsymbol{B} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top$. Then $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \boldsymbol{B}\boldsymbol{\varepsilon}$, and therefore

$$\mathrm{Var}\left(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) = \boldsymbol{B}\,\mathrm{diag}(v_1, \ldots, v_n)\,\boldsymbol{B}^\top.$$

The ordinary HC0 covariance estimator replaces v_i by the squared residual r_i^2, where $\boldsymbol{r} = (r_1, \ldots, r_n)^\top = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$.

However, r_i^2 is not an unbiased estimator of v_i. The next section gives an exact finite-sample correction.

3. Hadamard-corrected covariance estimator

Let $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top$ be the hat matrix and $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{H}$ be the residual-maker matrix. The least-squares residual vector is

$$\boldsymbol{r} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{M}\boldsymbol{Y}.$$

Because $\boldsymbol{M}\boldsymbol{X} = \boldsymbol{0}$, we have $\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\varepsilon}$.

Let m_{ij} denote the element of $\boldsymbol{M}$ at position (i, j). Then

$$r_i = \sum_{j=1}^{n} m_{ij}\, \varepsilon_j.$$

By conditional uncorrelatedness of the errors,

$$\mathrm{E}\left(r_i^2 \mid \boldsymbol{X} \right) = \sum_{j=1}^{n} m_{ij}^2\, v_j.$$

In vector form, $\mathrm{E}(\boldsymbol{r} \circ \boldsymbol{r} \mid \boldsymbol{X}) = (\boldsymbol{M} \circ \boldsymbol{M})\boldsymbol{v}$, where $\circ$ denotes the Hadamard, or elementwise, product, and $\boldsymbol{v} = (v_1, \ldots, v_n)^\top$.

Define $\boldsymbol{A} = \boldsymbol{M} \circ \boldsymbol{M}$.

If $\boldsymbol{A}$ is nonsingular, set $\widehat{\boldsymbol{v}}_H = \boldsymbol{A}^{-1}(\boldsymbol{r} \circ \boldsymbol{r})$. Then $\mathrm{E}(\widehat{\boldsymbol{v}}_H \mid \boldsymbol{X}) = \boldsymbol{v}$. Thus, $\widehat{\boldsymbol{v}}_H$ is a finite-sample unbiased estimator of the vector of individual Bernoulli variances. This immediately yields a finite-sample unbiased covariance estimator for $\widehat{\boldsymbol{\beta}}$:

$$\widehat{\mathrm{Var}}_H\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{B}\, \mathrm{diag}(\widehat{\boldsymbol{v}}_H)\, \boldsymbol{B}^\top.$$

The following theorem summarizes the result.

Theorem 1. Assume that the regression errors are conditionally uncorrelated given $\boldsymbol{X}$, that $\mathrm{E}(\boldsymbol{Y} \mid \boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\beta}_0$, that $\boldsymbol{X}$ has full column rank, and that $\boldsymbol{A} = \boldsymbol{M} \circ \boldsymbol{M}$ is nonsingular. Then

$$\mathrm{E}\left\{ \widehat{\mathrm{Var}}_H\left(\widehat{\boldsymbol{\beta}}\right) \mid \boldsymbol{X} \right\} = \mathrm{Var}\left(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X} \right).$$

Consequently, for any fixed contrast vector $\boldsymbol{c}$,

$$\mathrm{E}\left\{ \boldsymbol{c}^\top \widehat{\mathrm{Var}}_H\left(\widehat{\boldsymbol{\beta}}\right)\boldsymbol{c} \mid \boldsymbol{X} \right\} = \mathrm{Var}\left(\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} \mid \boldsymbol{X} \right).$$

Proof: Since $\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\varepsilon}$, the $\boldsymbol{i}$-th squared residual satisfies

$$\mathrm{E}\left(r_i^2 \mid \boldsymbol{X} \right) = \sum_{j=1}^{n} m_{ij}^2\, v_j.$$

Hence $\mathrm{E}(\boldsymbol{r} \circ \boldsymbol{r} \mid \boldsymbol{X}) = \boldsymbol{A}\boldsymbol{v}$. Because $\boldsymbol{A}$ is nonsingular,

$$\mathrm{E}\{ \widehat{\boldsymbol{v}}_H \mid \boldsymbol{X} \} = \mathrm{E}\{ \boldsymbol{A}^{-1}(\boldsymbol{r} \circ \boldsymbol{r}) \mid \boldsymbol{X} \} = \boldsymbol{v}.$$

Finally, $\widehat{\mathrm{Var}}_H(\widehat{\boldsymbol{\beta}}) = \boldsymbol{B}\mathrm{diag}(\widehat{\boldsymbol{v}}_H)\boldsymbol{B}^\top$ is linear in $\widehat{\boldsymbol{v}}_H$, and therefore its conditional expectation is

$$\boldsymbol{B}\mathrm{diag}(\boldsymbol{v})\boldsymbol{B}^\top = \mathrm{Var}(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X}).$$

This proves the theorem. □

Although Theorem 1 is stated conditionally on $\boldsymbol{X}$, it also implies marginal unbiasedness under a random-covariate design, provided that the assumptions of Theorem 1, including full column rank of $\boldsymbol{X}$ and nonsingularity of $\boldsymbol{A} = \boldsymbol{M} \circ \boldsymbol{M}$, hold almost surely and the relevant moments are finite. This follows because $\mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X}) = \boldsymbol{\beta}_0$, so that $\mathrm{Var}\{\mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X})\} = \boldsymbol{0}$ and hence $\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \mathrm{E}\{\mathrm{Var}(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X})\}$. If attention is restricted to an admissible design set determined solely by $\boldsymbol{X}$, the conclusion holds conditionally on that set.

4. Practical remarks

The theorem gives exact finite-sample unbiasedness for the covariance matrix, but several practical qualifications are important. First, the elements of $\widehat{\boldsymbol{v}}_H$ are not constrained to be nonnegative. Although the true Bernoulli variance satisfies $0 \le v_i \le 1/4$, an unbiased estimator of v_i need not satisfy these inequalities in every sample. As a result, $\widehat{\mathrm{Var}}_H(\widehat{\boldsymbol{\beta}})$ need not be positive semidefinite. This is not a contradiction; it is a consequence of insisting on exact unbiasedness for many nuisance variance components from a single realization of the response vector. Second, the estimator requires $\boldsymbol{A}$ to be nonsingular and reasonably well conditioned. When the design contains extreme leverage points or sparse or highly discrete covariate patterns, or when n is close to p, $\boldsymbol{A}$ can be singular or ill-conditioned. A general necessary condition for invertibility is $n \ge p + 1/2 + \sqrt{2p + 1/4}$, although this condition is not sufficient for every fixed design (Dobriban et al., 2018). In our implementation, numerical conditioning was assessed using the estimated reciprocal 1-norm condition number returned by the R function `rcond()`, and the Hadamard system was regarded as numerically evaluable when this quantity was at least 10^{-12}; this

threshold is a numerical safeguard rather than a theoretical condition for unbiasedness. Because $\boldsymbol{M}$ is symmetric and idempotent, $\boldsymbol{A} = \boldsymbol{M} \circ \boldsymbol{M}$ is positive semidefinite by the Schur product theorem and satisfies $\boldsymbol{A}\mathbf{1} = \mathbf{1} - \boldsymbol{h}$, where $\boldsymbol{h}$ is the leverage vector. These identities also clarify the relationship with HC2 and HC3: HC2 uses a row-sum-based diagonal approximation to the same system, and HC3 uses a diagonal (Jacobi-type) approximation, whereas HAD uses the full inverse $\boldsymbol{A}^{-1}$. When $\boldsymbol{A}$ is singular or numerically unstable, HAD should be regarded as non-evaluable rather than forced by inversion. Because implementation requires storing and solving an $n \times n$ linear system, computational and memory requirements may also become substantial for large n. This feature is also known in the broader literature on unbiased heteroskedastic variance estimation (Bera et al., 2002; Dobriban et al., 2018; Rao, 1970). Third, the finite-sample unbiasedness result depends on the correctness of the linear mean model. If $\mathrm{E}(\boldsymbol{Y} \mid \boldsymbol{X})$ is not exactly equal to $\boldsymbol{X}\boldsymbol{\beta}_0$, then the residual vector contains both random error and systematic lack of fit, and the preceding argument no longer applies directly. This limitation is shared with the usual interpretation of the least-squares estimator as a model-based risk-difference estimator. Fourth, the term "exact" must be interpreted carefully. The proposed covariance estimator is exact in the sense that its conditional expectation equals the true covariance matrix of $\widehat{\boldsymbol{\beta}}$ in finite samples. This does not imply that the usual Wald interval,

$$\boldsymbol{c}^{\top}\widehat{\boldsymbol{\beta}} \pm z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\boldsymbol{c}^{\top}\widehat{\mathrm{Var}}_H\left(\widehat{\boldsymbol{\beta}}\right)\boldsymbol{c}},$$

has exact finite-sample coverage. Exact coverage would require an exact pivotal distribution or an exact inversion argument, which is not available here in general. The proposed estimator is therefore best regarded as an exact finite-sample unbiased covariance estimator, not as a finite-sample exact confidence-interval procedure.

For applied use, stabilized versions may be considered. Let $\hat{v}_{H,i}$ denote the i-th component of $\widehat{\boldsymbol{v}}_H$. One may define the nonnegative truncated estimator componentwise by

$$\hat{v}_{H,+,i} = \max(\hat{v}_{H,i}, 0)$$

or, using the Bernoulli variance bound, the range-truncated estimator componentwise by

$$\hat{v}_{H,[0,1/4],i} = \min\{\max(\hat{v}_{H,i}, 0), 1/4\}.$$

The corresponding vectors are denoted by $\widehat{\boldsymbol{v}}_{H,+}$ and $\widehat{\boldsymbol{v}}_{H,[0,1/4]}$, respectively, and can be substituted for $\widehat{\boldsymbol{v}}_H$ in the covariance estimator. These versions may improve numerical stability and guarantee nonnegative individual variance estimates. However, because the truncation is nonlinear, they no longer preserve the exact finite-sample unbiasedness property. They should therefore be viewed as pragmatic sensitivity analyses rather than as the main theoretical proposal. The truncation to [0, 1/4] is specific to Bernoulli outcomes and is not generally appropriate for count outcomes or incidence rate difference estimation.

The same argument applies to the weighted least-squares estimator for incidence rate differences proposed by Xu et al. (2010), provided that the analysis is based on independent individual-level observations and fixed person-time. Let Y_i be the event count, t_i the follow-up time, and $\lambda_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$ the incidence rate. The additive rate model $\mathrm{E}(Y_i|\boldsymbol{X}_i, t_i) = t_i \boldsymbol{x}_i^\top \boldsymbol{\beta}$ can be written as an ordinary least-squares regression of $Y_i/\sqrt{t_i}$ on $\sqrt{t_i}\boldsymbol{x}_i$. Thus, applying the Hadamard correction to the residual-maker matrix from this transformed regression yields a finite-sample unbiased estimator of the transformed error variances, and hence an analogous covariance correction for the incidence rate difference estimator. Extensions to recurrent-event data with within-subject dependence would require a cluster-level or block-Hadamard version and are left for future work.

The computation of the unbiased covariance estimator for modified least-squares regression, together with the corresponding confidence intervals, is implemented in the `rqlm` package available from CRAN (https://doi.org/10.32614/CRAN.package.rqlm).

5. Simulation study

To evaluate the finite-sample performance of the proposed estimator, we conducted a simulation study. The simulation design was based on the practical settings considered by Noma and Gosho (2025). We generated four binary covariates. The primary exposure x_{i1} was binary with marginal probability 0.10 or 0.20. The binary confounder x_{i2} had marginal probability 0.773 and was associated with x_{i1}; the joint distribution of x_{i1} and x_{i2} was chosen so that the odds ratio for their 2 × 2 table was 5 or 10. The covariate x_{i3} was binary with probability 0.455, and x_{i4} was also binary with probability 0.50.

Conditional on the covariates, the outcome Y_i was generated from a Bernoulli distribution with probability

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

The intercept β_0 was calibrated to yield a marginal event probability of 0.20 or 0.40. The remaining regression coefficients ($\beta_1, \beta_2, \beta_3, \beta_4$) were fixed at (0.116, -0.041, 0.0037, 0.023). The complete covariate support was checked in advance to ensure that all Bernoulli probabilities were strictly between 0 and 1. Across the nonzero-effect scenarios, the true probabilities ranged from 0.154 to 0.550.

We considered sample sizes n = 40, 70, and 100 to represent finite-sample settings in which conventional heteroskedasticity-robust variance estimators may perform poorly. For each of the 24 scenarios, 10,000 Monte Carlo replications were attempted. Corresponding null simulations were conducted with $\beta_1 = 0$ to estimate Type I error, with the intercept recalibrated to preserve the target marginal event probability. For each full-rank design matrix, $\boldsymbol{B}\mathrm{diag}\{p_i(1-p_i)\}_{i=1}^{n}\boldsymbol{B}^{\top}$ was used as the analytic true conditional covariance matrix. For each simulated data set, we fitted modified least-squares regression with an identity link and estimated the variance of the exposure coefficient β_1. As comparators, we used the HC0, HC1, HC2, and HC3 heteroskedasticity-robust covariance estimators as well as Morel–Bokossa–Neerchal (MBN), Gosho–Sato–Takeuchi (GST), and Wang–Long (WL) small-sample corrections. The covariance estimators were calculated directly from their formulas to ensure consistent handling of numerical

issues across methods. We also evaluated the proposed Hadamard estimator (HAD) and two stabilized variants, HAD0 and HAD025, as sensitivity analyses. For HAD, the observation-level variance components were estimated as $\boldsymbol{A}^{-1}(\boldsymbol{r} \circ \boldsymbol{r})$, where $\boldsymbol{A} = (\boldsymbol{I} - \boldsymbol{H}) \circ (\boldsymbol{I} - \boldsymbol{H})$, $\boldsymbol{H}$ is the hat matrix, and $\boldsymbol{r}$ is the residual vector. HAD0 replaced negative estimated variance components by zero, whereas HAD025 truncated each component to the interval [0, 0.25].

The common evaluable design set comprised full-rank designs for which the Hadamard matrix $\boldsymbol{A} = \boldsymbol{M} \circ \boldsymbol{M}$ satisfied the prespecified numerical conditioning criterion. Bias and root mean squared error (RMSE) were evaluated among method-computable replications within this common set. Wald coverage and Type I error were evaluated among replications for which the target variance was positive and the corresponding confidence interval was constructible. Rank-deficient designs and method-specific numerical non-evaluability were classified as computational failures, whereas negative HAD components, non-positive-semidefinite covariance matrices, and nonpositive target variances were recorded separately as diagnostic outcomes. Performance comparisons were based on the common evaluable design set, while computational feasibility was summarized using all attempted datasets. In addition, we performed a fixed-covariate Monte Carlo experiment using prespecified regular and challenging designs; the exact fixed design matrices and corresponding probability vectors are provided in the accompanying R code. For each design, the covariates and analytic true variance were held fixed and 100,000 Bernoulli outcome vectors were generated to compare the Monte Carlo mean variance estimates of HAD and HC2 with the analytic target.

HAD closely tracked the analytic true variance throughout the simulations (Table 1 and Supplementary Table S1). Its bias ranged from −0.0002 to 0.0003 for $n = 40$, from −0.0003 to 0.0001 for $n = 70$, and from −0.0001 to approximately zero for $n = 100$. HC2 also showed small bias, whereas HC0 and HC1 generally underestimated the variance and HC3 generally overestimated it. Thus, HAD provided particularly stable control of systematic variance-estimation bias across the range of settings considered.

Table 1. Summary of simulation studies

n	Variance estimator	Mean true variance, range	Mean estimated variance, range	Bias, range	95% CI coverage, range	Computable analyses (%)
40	HAD	0.0354–0.0713	0.0354–0.0712	-0.0002 to 0.0003	0.762–0.916	88.2
	HC0	0.0354–0.0713	0.0283–0.0520	-0.0193 to -0.0072	0.741–0.898	99.3
	HC1	0.0354–0.0713	0.0323–0.0594	-0.0119 to -0.0031	0.754–0.914	99.3
	HC2	0.0354–0.0713	0.0350–0.0709	-0.0008 to 0.0002	0.764–0.919	96.9
	HC3	0.0354–0.0713	0.0438–0.0988	0.0080 to 0.0279	0.784–0.934	96.9
	MBN	0.0354–0.0713	0.0366–0.0692	-0.0039 to 0.0032	0.792–0.931	99.3
	GST	0.0354–0.0713	0.0295–0.0688	-0.0117 to -0.0015	0.912–0.941	99.3
	WL	0.0354–0.0713	0.0343–0.0795	-0.0040 to 0.0081	0.933–0.958	96.9
70	HAD	0.0190–0.0456	0.0191–0.0457	-0.0003 to 0.0001	0.859–0.931	98.8
	HC0	0.0190–0.0456	0.0169–0.0371	-0.0089 to -0.0021	0.846–0.919	100.0
	HC1	0.0190–0.0456	0.0182–0.0399	-0.0061 to -0.0008	0.853–0.928	100.0
	HC2	0.0190–0.0456	0.0189–0.0456	-0.0004 to 0.0000	0.859–0.931	99.9
	HC3	0.0190–0.0456	0.0212–0.0568	0.0021 to 0.0112	0.868–0.944	99.9
	MBN	0.0190–0.0456	0.0195–0.0433	-0.0029 to 0.0010	0.858–0.939	100.0
	GST	0.0190–0.0456	0.0158–0.0438	-0.0080 to -0.0008	0.916–0.946	100.0
	WL	0.0190–0.0456	0.0172–0.0474	-0.0054 to 0.0018	0.927–0.953	99.9
100	HAD	0.0131–0.0315	0.0130–0.0315	-0.0001 to 0.0000	0.901–0.942	99.9
	HC0	0.0131–0.0315	0.0120–0.0273	-0.0042 to -0.0010	0.888–0.934	100.0
	HC1	0.0131–0.0315	0.0127–0.0287	-0.0028 to -0.0004	0.894–0.940	100.0
	HC2	0.0131–0.0315	0.0130–0.0315	-0.0002 to -0.0000	0.902–0.942	100.0
	HC3	0.0131–0.0315	0.0140–0.0364	0.0009 to 0.0050	0.911–0.950	100.0
	MBN	0.0131–0.0315	0.0133–0.0303	-0.0014 to 0.0005	0.900–0.945	100.0
	GST	0.0131–0.0315	0.0108–0.0303	-0.0056 to -0.0005	0.914–0.946	100.0
	WL	0.0131–0.0315	0.0114–0.0320	-0.0044 to 0.0005	0.922–0.951	100.0

Note. Values for the mean true variance, mean estimated variance, bias, and confidence-interval coverage are shown as ranges across the eight simulation scenarios with a nonzero exposure effect within each sample size. Coverage is based on two-sided 95% Wald confidence intervals using the standard normal critical value and is evaluated among constructible intervals. Computable analyses are expressed as the percentage of all attempted datasets. Detailed causes of non-computability are reported in Supplementary Table S3, Panel A.

Confidence-interval coverage improved with increasing sample size and was broadly similar for HAD and HC2. For HAD, coverage lay in the ranges 0.762–0.916, 0.859–0.931, and 0.901–0.942 for n = 40, 70, and 100, respectively

(Table 1). Despite the small bias of the HAD covariance estimator, HAD-based Wald inference was substantially anti-conservative in several sparse settings. Among constructible tests, Type I error reached 24.8% at the nominal two-sided 5% level (Supplementary Table S2), whereas GST and WL generally showed Type I error rates closer to 0.05. Thus, unbiased covariance estimation did not ensure adequate finite-sample calibration of Wald tests. The computational feasibility of HAD increased from 88.2% at $n = 40$ to 98.8% at $n = 70$ and 99.9% at $n = 100$, with most HAD failures attributable to numerical conditioning of the Hadamard system (Supplementary Table S3). Although non-positive-semidefinite covariance estimates occurred in some samples, a nonpositive target variance was uncommon, with a decrease from 0.202% at n = 40 to 0.002% at n = 100. Observation-level variance estimates outside the Bernoulli variance range [0, 0.25] were nevertheless frequent. Among computable HAD analyses pooled across the nonzero-effect and null simulations, at least one negative component occurred in 85.2%, 58.8%, and 38.4% of datasets for n = 40, 70, and 100, respectively; the percentage with at least one component exceeding 0.25 rounded to 100.0% at each sample size (Supplementary Table S3, Panel B).

The fixed-covariate experiment further supported the finite-sample accuracy of HAD (Table 2). In both designs, the Monte Carlo mean of the HAD variance estimate was closer to the analytic true variance than that of HC2. In the challenging design, for example, the true variance was 0.04982, compared with mean estimates of 0.04967 for HAD and 0.04924 for HC2; the corresponding RMSEs were 0.02377 and 0.02237. Thus, the reduction in systematic bias with HAD was accompanied by only a modest increase in RMSE. The sensitivity analyses also showed that truncation altered this property: HAD0 introduced modest upward bias, whereas HAD025 produced appreciable downward bias, particularly at smaller sample sizes (Supplementary Table S4). For HAD025, the reduction caused by capping numerous large positive component estimates at 0.25 outweighed the upward effect of replacing negative component estimates by zero.

Table 2. Monte Carlo comparison of HAD and HC2 with the analytic true variance under fixed covariate designs

Design	Method	True variance	Mean estimated variance	Bias (MCSE)	RMSE
Regular	HAD	0.01083	0.01082	-0.000005 (0.000006)	0.00184
	HC2	0.01083	0.01079	-0.000035 (0.000006)	0.00182
Challenging	HAD	0.04982	0.04967	-0.000149 (0.000075)	0.02377
	HC2	0.04982	0.04924	-0.000583 (0.000071)	0.02237

Note. The covariate design matrix was held fixed and only the Bernoulli outcomes were repeatedly generated. True variance denotes the analytic conditional variance for each fixed design. Mean estimated variance is the Monte Carlo mean from 100,000 outcome replications. Bias is the mean estimated variance minus the analytic true variance. MCSE denotes the Monte Carlo standard error of the estimated bias.

R code for the primary simulation study, the fixed-design experiment, and the sensitivity analysis using t_{n-p} critical values is included in the supplementary R code archive, which is available upon request by email to the corresponding author.

6. Application to a randomized respiratory trial

We illustrated the methods using the publicly available respiratory data distributed with the R package `geepack` (Halekoh, Højsgaard and Yan, 2006). The data originate from a randomized, two-center clinical trial of an active treatment versus placebo in patients with respiratory illness (Davis, 1991). Respiratory status was recorded as good or poor at baseline and at four scheduled visits during treatment. The trial included 111 participants, 54 assigned to active treatment and 57 assigned to placebo.

The original data have a longitudinal structure. Because the theoretical development in this paper assumes independent individual-level observations, we prespecified respiratory status at the fourth and final scheduled visit as the binary endpoint and used one record per participant. The analysis is therefore an

illustration of adjusted final-visit risk-difference estimation rather than a replacement for a full longitudinal analysis.

Let $Y_i = 1$ denote good respiratory status at visit 4 and $Y_i = 0$ denote poor status. We fitted the model

$$\mathrm{E}(Y_i \mid \boldsymbol{X}) = \beta_0 + \beta_1 \mathrm{Active}_i + \beta_2 \mathrm{Center2}_i + \beta_3 \mathrm{Age10}_i + \beta_4 \mathrm{BaselineGood}_i,$$

where $\mathrm{Active}_i = 1$ for active treatment and zero for placebo, $\mathrm{Center2}_i = 1$ for center 2, age was centered at the sample mean and divided by 10, and $\mathrm{BaselineGood}_i = 1$ for good baseline respiratory status. The treatment coefficient β_1 is the adjusted final-visit risk difference between active treatment and placebo.

We calculated HAD, HAD0, HAD025, and HC0–HC3 covariance estimates and constructed 95% Wald intervals. We also examined the leverage values, the condition number of $\boldsymbol{A}$, the range of the fitted probabilities, and the observation-specific variance estimates. A sensitivity analysis using t_{n-p} rather than standard normal critical values yielded slightly wider confidence intervals. The HAD interval remained above zero, whereas the HC3 interval included zero. Standard normal critical values were retained for the primary analysis.

At the final visit, good respiratory status was observed in 33 of 54 participants assigned to active treatment (61.1%) and 25 of 57 assigned to placebo (43.9%). The unadjusted risk difference was therefore 0.173. After adjustment for center, age, and baseline respiratory status, the estimated treatment risk difference was $\hat{\beta}_1 = 0.170$. The maximum leverage was 0.089, indicating that no extreme leverage was present. The 2-norm condition number of A was 1.160, indicating that the Hadamard-square system was numerically well conditioned.

The fitted risks ranged from 0.013 to 1.021. The small excursion above one illustrates a familiar feature of linear probability models, although there was no substantial extrapolation in this application. This out-of-range fitted value is a finite-sample feature of the estimated linear probability model and should be distinguished from the assumption in Theorem 1 that the true conditional means are valid Bernoulli probabilities in $[0, 1]$. The untruncated observation-specific variance estimates ranged from -0.015 to 0.988; seven were negative and 28

exceeded 0.25. Despite these component-level values, the HAD covariance matrix was positive definite and the estimated treatment variance was positive.

Table 3 presents the estimated standard errors and 95% confidence intervals. HAD and HAD0 produced almost identical results. HC1 and HC2 were also nearly identical to HAD, whereas HC0 gave a slightly smaller standard error and HC3 a slightly larger one. HAD025 produced a substantially smaller standard error because it truncated 28 observation-specific components above 0.25. Given its systematic downward bias in the simulation study, the narrower HAD025 interval should not be interpreted as evidence of superior precision.

The analysis suggests an approximately 17-percentage-point higher probability of good respiratory status under active treatment at the final visit, after adjustment for the prespecified covariates. The lower limits under HAD and the conventional HC estimators were close to zero, illustrating that interpretation can be sensitive to finite-sample standard-error estimation even when the point estimate is unchanged.

Table 3. Adjusted final-visit risk difference in the respiratory trial

Variance estimator	Adjusted risk difference	Standard error	95% confidence interval
HAD	0.170	0.084	(0.005, 0.335)
HAD0	0.170	0.084	(0.005, 0.335)
HAD025	0.170	0.069	(0.035, 0.306)
HC0	0.170	0.082	(0.009, 0.332)
HC1	0.170	0.084	(0.005, 0.335)
HC2	0.170	0.084	(0.005, 0.335)
HC3	0.170	0.086	(0.001, 0.339)

Note. The risk difference compares active treatment with placebo and is adjusted for study center, age in 10-year units, and baseline respiratory status.

R code reproducing all covariance estimators in Table 3 and the sensitivity analysis using t_{n-p} critical values is included in the supplementary R code archive, which is available upon request by email to the corresponding author.

7. Concluding remarks

Modified least-squares regression is appealing because it estimates covariate-adjusted risk differences by ordinary least squares applied to a binary response (Cheung, 2007). Under a correctly specified linear mean model and fixed covariates, this simplicity yields a useful finite-sample property: the covariance matrix of the estimator can be estimated without finite-sample bias by correcting the squared residuals through the Hadamard square of the residual-maker matrix. The proposed estimator therefore provides a finite-sample counterpart to the usual heteroskedasticity-consistent covariance estimators, which are primarily justified by asymptotic arguments (Eicker, 1963; Huber, 1967; MacKinnon and White, 1985; White, 1980).

The result connects modified least-squares regression with classical unbiased variance estimation in heteroskedastic linear models (Hartley et al., 1969; Rao, 1970; Bera et al., 2002; Dobriban et al., 2018). Its limitations are also clear: the linear risk model must be correctly specified, $\boldsymbol{M} \circ \boldsymbol{M}$ must be nonsingular, individual variance estimates may be negative, and the resulting covariance matrix need not be positive semidefinite. Moreover, finite-sample unbiased covariance estimation does not imply exact finite-sample confidence-interval coverage. The proposed method should therefore be viewed as a simple finite-sample covariance correction for modified least-squares regression.

Acknowledgements

The authors were supported by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (grant numbers: JP23K24811, JP23K11931, and JP26K02873). The authors also thank an artificial intelligence–based language model (ChatGPT, OpenAI) for assistance with English language editing and stylistic refinement of the manuscript. The authors take full responsibility for the content, interpretation, and conclusions presented in this article.

Conflict of interest: The authors declare no conflicts of interest.

Data availability statement

The respiratory data analyzed in this study are publicly available in the `geepack` R package. The R code for the simulations and application is available upon request by email to the corresponding author.

REFERENCES

Altman D.G., Deeks J.J., Sackett D.L. (1998): Odds Ratios Should Be Avoided When Events Are Common. BMJ 317(7168): 1318.

Bera A.K., Suprayitno T., Premaratne G. (2002): On Some Heteroskedasticity-Robust Estimators of Variance-Covariance Matrix of the Least-Squares Estimators. Journal of Statistical Planning and Inference 108(1-2): 121-136.

Chesher A., Austin G. (1991): The Finite-Sample Distributions of Heteroskedasticity Robust Wald Statistics. Journal of Econometrics 47(1): 153-173.

Chesher A., Jewitt I. (1987): The Bias of a Heteroskedasticity Consistent Covariance Matrix Estimator. Econometrica 55(5): 1217-1222.

Cheung Y. B. (2007): A Modified Least-Squares Regression Approach to the Estimation of Risk Difference. American Journal of Epidemiology 166(11): 1337-1344.

Cribari-Neto F. (2004): Asymptotic Inference under Heteroskedasticity of Unknown Form. Computational Statistics & Data Analysis 45(2): 215-233.

Cribari-Neto F., Souza T.C., Vasconcellos K.L.P. (2007): Inference under Heteroskedasticity and Leveraged Data. Communications in Statistics—Theory and Methods 36(10): 1877-1888.

Davis C.S. (1991): Semi-Parametric and Non-Parametric Methods for the Analysis of Repeated Measurements with Applications to Clinical Trials. Statistics in Medicine 10(12): 1959-1980.

Dobriban E., Su W.J., Yang Y., Zhang Z. (2018): Robust Inference under Heteroskedasticity Via the Hadamard Estimator. arXiv: 1807.00347.

Eicker F. (1963): Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions. Annals of Mathematical Statistics 34(2): 447–456.

Eicker F. (1967). Limit Theorems for Regressions with Unequal and Dependent Errors. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, L. M. Le Cam, and J. Neyman (eds), 59-82. Berkeley: University of California Press.

Fay M.P., Graubard B.I. (2001): Small-Sample Adjustments for Wald-Type Tests Using Sandwich Estimators. Biometrics 57(4): 1198-1206.

Godambe V.P., Heyde C.C. (1987): Quasi-Likelihood and Optimal Estimation. International Statistical Review 55(3): 231-244.

Gosho M., Sato Y., Takeuchi H. (2014): Robust Covariance Estimator for Small-Sample Adjustment in the Generalized Estimating Equations: A Simulation Study. Science Journal of Applied Mathematics and Statistics 2(1): 20-25.

Halekoh U., Højsgaard S., Yan J. (2006): The R Package Geepack for Generalized Estimating Equations. Journal of Statistical Software 15(2): 1-11.

Hartley H.O., Rao J.N.K., Kiefer G. (1969): Variance Estimation with One Unit Per Stratum. Journal of the American Statistical Association 64(327): 841-851.

Hinkley D.V. (1977): Jackknifing in Unbalanced Situations. Technometrics 19(3): 285-292.

Hopewell S., Chan A.W., Collins G.S., Hrobjartsson A., Moher D., Schulz K.F., Tunn R., Aggarwal R., Berkwits M., Berlin J.A., Bhandari N., Butcher N.J., Campbell M.K., Chidebe R.C.W., Elbourne D., Farmer A., Fergusson D.A., Golub R.M., Goodman S.N., Hoffmann T.C., Ioannidis J.P.A., Kahan B.C., Knowles R.L., Lamb S.E., Lewis S., Loder E., Offringa M., Ravaud P., Richards D.P., Rockhold F.W., Schriger D.L., Siegfried N.L., Staniszewska S., Taylor R.S., Thabane L., Torgerson D., Vohra S., White I.R., Boutron I. (2025): CONSORT 2025 Statement: Updated Guideline for Reporting Randomized Trials. Nature Medicine. 31(6):1776-1783.

Huber P.J. (1967). The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, L. Le Cam, and J. Neyman (eds), 221–233. Berkeley: University of California Press.

Long J.S., Ervin L.H. (2000): Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. American Statistician 54(3): 217-224.

MacKinnon J.G., White H. (1985): Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties. Journal of Econometrics 29(3): 305-325.

Mancl L.A., DeRouen T.A. (2001): A Covariance Estimator for GEE with Improved Small-Sample Properties. Biometrics 57(1): 126-134.

McNutt L.A., Wu C., Xue X., Hafner J.P. (2003): Estimating the Relative Risk in Cohort Studies and Clinical Trials of Common Outcomes. American Journal of Epidemiology 157(10): 940-943.

Morel J.G., Bokossa M., Neerchal N. (2003): Small Sample Correction for the Variance of GEE Estimators. Biometrical Journal 45(4): 395-409.

Nelder J.A., Wedderburn R.W.M. (1972): Generalized Linear Models. Journal of the Royal Statistical Society, Series A 135(3): 370-384.

Noma H., Gosho M. (2025): Finite-Sample Improved Confidence Intervals Based on the Estimating Equation Theory for the Modified Poisson and Least-Squares Regressions. Epidemiologic Methods 14(1): 20240030.

Noma H., Sunada H., Gosho M. (2025): Quasi-Likelihood Ratio Tests and the Bartlett-Type Correction for Improved Inferences of the Modified Poisson and Least-Squares Regressions for Binary Outcomes. Statistica Neerlandica 79(3): e70012.

Nurminen M. (1995): To Use or Not to Use the Odds Ratio in Epidemiologic Analyses. European Journal of Epidemiology 11(4): 365-371.

Pan W., Wall M.M. (2002): Small-Sample Adjustments in Using the Sandwich Variance Estimator in Generalized Estimating Equations. Statistics in Medicine 21(10): 1429-1441.

Rao C.R. (1970): Estimation of Heteroscedastic Variances in Linear Models. Journal of the American Statistical Association 65: 161-172.

Rothman K.J., Greenland S., Lash T.L. (2008). Modern Epidemiology, 3rd edition. Lippincott Williams & Wilkins, Philadelphia.

Thompson J., Watson S.I., Middleton L., Hemming K. (2025): Estimating Relative Risks and Risk Differences in Randomised Controlled Trials: A Systematic Review of Current Practice. Trials 26(1): 1.

Uno S., Noma H., Gosho M. (2024): Firth-Type Penalized Methods of the Modified Poisson and Least-Squares Regression Analyses for Binary Outcomes. Biometrical Journal 66(7): e202400004.

Wallenstein S., Bodian C. (1987): Epidemiologic Programs for Computers and Calculators. Inferences on Odds Ratios, Relative Risks, and Risk Differences Based on Standard Regression Programs. American Journal of Epidemiology 126(2): 346-355.

Wang M., Long Q. (2011): Modified Robust Variance Estimator for Generalized Estimating Equations with Improved Small-Sample Performance. Statistics in Medicine 30(11): 1278-1291.

Wedderburn R.W.M. (1974): Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss–Newton Method. Biometrika 61(3): 439-447.

White H. (1980): A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica 48(4): 817-838.

Xu Y., Cheung Y.B., Lam K.F., Tan S.H., Milligan P. (2010): A Simple Approach to the Estimation of Incidence Rate Difference. American Journal of Epidemiology 172(3): 334-343.

Supplementary Material

Exact unbiased covariance estimation for modified least-squares regression with binary outcomes

Takahiro Kitano[1], Hisashi Noma[1,2]

[1] The Graduate Institute for Advanced Studies, The Graduate University for Advanced Studies, Tokyo, Japan
[2] Department of Interdisciplinary Statistical Mathematics, The Institute of Statistical Mathematics, Tokyo, Japan, e-mail: noma@ism.ac.jp

Overview

This supplementary material provides detailed results from the simulation study described in section 5 of the main manuscript. The primary simulation comprised 24 scenarios crossing marginal event probabilities of 0.20 and 0.40, exposure prevalences of 0.10 and 0.20, exposure–confounder odds ratios of 5 and 10, and sample sizes of 40, 70, and 100. For each scenario, 10,000 Monte Carlo replications were performed under the prespecified nonzero exposure effect. A separate set of 10,000 replications per scenario was performed under the null hypothesis of no exposure effect, with the intercept recalibrated to preserve the target marginal event probability.

Supplementary Table S1 reports the complete scenario-specific results under the nonzero exposure effect for the proposed Hadamard estimator (HAD), the conventional HC0–HC3 heteroskedasticity-robust covariance estimators, and the MBN, GST, and WL small-sample covariance corrections. For each method, the table reports the mean analytic true variance of the exposure coefficient β1, the mean estimated variance, bias, root mean squared error (RMSE), 95% Wald confidence-interval coverage, and mean confidence-interval width.

Supplementary Table S2 reports the corresponding scenario-specific Type I error rates under the null hypothesis $\beta_1 = 0$ for the same variance estimators. Type I error was evaluated using two-sided Wald tests based on the standard normal critical value.

Supplementary Table S3 summarizes computational feasibility and diagnostic outcomes. Panel A reports the percentages of computable analyses and the principal causes of non-computability, including rank-deficient design matrices, numerical conditioning failure of the Hadamard system, and leverage-related non-evaluability. These percentages use all attempted datasets as the denominator and are pooled across the nonzero-effect and null simulations. Panel B reports diagnostic outcomes for the untruncated HAD estimator among computable HAD analyses, including negative estimated variance components, components exceeding 0.25, non-positive-semidefinite covariance matrices, and nonpositive estimated variances for the target coefficient. These diagnostic outcomes were recorded separately and were not themselves classified as computational failures.

Supplementary Table S4 provides a sensitivity analysis of the stabilized Hadamard estimators. HAD0 replaces negative estimated variance components by zero, whereas HAD025 truncates the estimated components to the interval [0, 0.25]. For HAD, HAD0, and HAD025, the table summarizes the ranges of the mean analytic true variance, mean estimated variance, bias, and RMSE across the eight simulation scenarios within each sample size.

The common design-evaluable set comprised full-rank designs for which the Hadamard matrix $\boldsymbol{A}$ satisfied the prespecified numerical conditioning criterion. Performance comparisons in Supplementary Tables S1 and S2 were based on this set, with method-specific summaries calculated when the corresponding covariance estimator was computable. Computational feasibility in Supplementary Table S3 was instead evaluated using all attempted datasets. Wald confidence-interval coverage and Type I error were evaluated only when a finite, positive variance estimate was available and the corresponding confidence interval was constructible.

HAD denotes the proposed untruncated Hadamard estimator. HC0–HC3 denote the conventional heteroskedasticity-robust covariance estimators. MBN, GST, and WL denote the Morel–Bokossa–Neerchal, Gosho–Sato–Takeuchi, and Wang–Long small-sample covariance corrections, respectively. HAD0 and HAD025 are stabilized versions of HAD used in the sensitivity analysis.

R code and software information

The accompanying R code archive contains `Rcode_Simulation.R`, covering the primary simulations, the fixed-design experiment in Table 2, and the t_{n-p}-based simulation sensitivity analysis, and `Rcode_Application.R`, covering all covariance estimators in Table 3 and the corresponding application sensitivity analysis. Reported software-version information is documented in the script headers; the application information refers to the local reference environment supplied with the code. The R code is available upon request by email to the corresponding author.

Definitions used in Supplementary Tables S1–S4

Analytic true variance = the conditional variance of the estimated exposure coefficient calculated from the simulated design matrix and the true Bernoulli variances.

Bias = mean estimated variance − mean analytic true variance, evaluated over the relevant method-computable replications within the common design-evaluable set.

RMSE = the root mean squared error of the estimated variance relative to the replication-specific analytic true variance.

95% CI coverage = the proportion of constructible two-sided 95% Wald confidence intervals containing the true exposure coefficient.

Type I error = the proportion of constructible two-sided Wald tests rejecting the null hypothesis $\beta_1 = 0$ when the null hypothesis is true.

Computable analyses (%) = the percentage of all attempted datasets for which the corresponding covariance estimator could be numerically evaluated.

Supplementary Table S1. Complete results from simulations with a nonzero exposure effect

Scenario	n	Event proba-bility	Exposure prevalence	OR	Method	Mean true variance	Mean estimated variance	Bias	RMSE	95% CI coverage	Mean CI width
1	40	0.2	0.1	5	HAD	0.0566	0.0567	0.0000	0.0380	0.771	0.862
1	40	0.2	0.1	5	HC0	0.0566	0.0405	-0.0161	0.0293	0.750	0.744
1	40	0.2	0.1	5	HC1	0.0566	0.0463	-0.0103	0.0293	0.761	0.795
1	40	0.2	0.1	5	HC2	0.0566	0.0559	-0.0007	0.0350	0.772	0.864
1	40	0.2	0.1	5	HC3	0.0566	0.0785	0.0218	0.0592	0.793	1.010
1	40	0.2	0.1	5	MBN	0.0566	0.0528	-0.0039	0.0280	0.799	0.861
1	40	0.2	0.1	5	GST	0.0566	0.0449	-0.0117	0.0173	0.926	0.816
1	40	0.2	0.1	5	WL	0.0566	0.0527	-0.0040	0.0149	0.946	0.884
2	70	0.2	0.1	5	HAD	0.0370	0.0369	-0.0001	0.0204	0.859	0.711
2	70	0.2	0.1	5	HC0	0.0370	0.0295	-0.0075	0.0171	0.846	0.644
2	70	0.2	0.1	5	HC1	0.0370	0.0317	-0.0052	0.0170	0.853	0.668
2	70	0.2	0.1	5	HC2	0.0370	0.0366	-0.0004	0.0196	0.860	0.711
2	70	0.2	0.1	5	HC3	0.0370	0.0459	0.0089	0.0296	0.868	0.788
2	70	0.2	0.1	5	MBN	0.0370	0.0340	-0.0029	0.0164	0.859	0.696

Scenario	n	Event proba-bility	Exposure prevalence	OR	Method	Mean true variance	Mean estimated variance	Bias	RMSE	95% CI coverage	Mean CI width
2	70	0.2	0.1	5	GST	0.0370	0.0289	-0.0080	0.0105	0.918	0.656
2	70	0.2	0.1	5	WL	0.0370	0.0316	-0.0054	0.0086	0.930	0.685
3	100	0.2	0.1	5	HAD	0.0256	0.0257	0.0000	0.0108	0.902	0.606
3	100	0.2	0.1	5	HC0	0.0256	0.0220	-0.0036	0.0095	0.889	0.565
3	100	0.2	0.1	5	HC1	0.0256	0.0232	-0.0025	0.0095	0.895	0.579
3	100	0.2	0.1	5	HC2	0.0256	0.0255	-0.0001	0.0105	0.903	0.605
3	100	0.2	0.1	5	HC3	0.0256	0.0297	0.0041	0.0146	0.913	0.649
3	100	0.2	0.1	5	MBN	0.0256	0.0243	-0.0014	0.0092	0.901	0.594
3	100	0.2	0.1	5	GST	0.0256	0.0200	-0.0056	0.0069	0.914	0.547
3	100	0.2	0.1	5	WL	0.0256	0.0213	-0.0044	0.0059	0.922	0.564
4	40	0.2	0.1	10	HAD	0.0570	0.0569	-0.0001	0.0383	0.762	0.863
4	40	0.2	0.1	10	HC0	0.0570	0.0408	-0.0162	0.0297	0.741	0.746
4	40	0.2	0.1	10	HC1	0.0570	0.0467	-0.0104	0.0297	0.754	0.798
4	40	0.2	0.1	10	HC2	0.0570	0.0562	-0.0008	0.0354	0.764	0.866
4	40	0.2	0.1	10	HC3	0.0570	0.0786	0.0216	0.0593	0.784	1.010
4	40	0.2	0.1	10	MBN	0.0570	0.0532	-0.0038	0.0284	0.792	0.865
4	40	0.2	0.1	10	GST	0.0570	0.0456	-0.0114	0.0173	0.922	0.823
4	40	0.2	0.1	10	WL	0.0570	0.0534	-0.0036	0.0149	0.945	0.890
5	70	0.2	0.1	10	HAD	0.0369	0.0368	-0.0001	0.0205	0.860	0.710
5	70	0.2	0.1	10	HC0	0.0369	0.0295	-0.0074	0.0171	0.846	0.643
5	70	0.2	0.1	10	HC1	0.0369	0.0318	-0.0052	0.0171	0.853	0.668
5	70	0.2	0.1	10	HC2	0.0369	0.0365	-0.0004	0.0197	0.859	0.710
5	70	0.2	0.1	10	HC3	0.0369	0.0457	0.0088	0.0299	0.868	0.785
5	70	0.2	0.1	10	MBN	0.0369	0.0340	-0.0029	0.0165	0.858	0.696
5	70	0.2	0.1	10	GST	0.0369	0.0289	-0.0080	0.0105	0.923	0.656
5	70	0.2	0.1	10	WL	0.0369	0.0315	-0.0054	0.0087	0.936	0.685
6	100	0.2	0.1	10	HAD	0.0256	0.0255	-0.0001	0.0110	0.901	0.603
6	100	0.2	0.1	10	HC0	0.0256	0.0219	-0.0036	0.0098	0.888	0.564
6	100	0.2	0.1	10	HC1	0.0256	0.0231	-0.0025	0.0098	0.894	0.578
6	100	0.2	0.1	10	HC2	0.0256	0.0254	-0.0002	0.0107	0.902	0.603
6	100	0.2	0.1	10	HC3	0.0256	0.0295	0.0039	0.0147	0.911	0.646
6	100	0.2	0.1	10	MBN	0.0256	0.0242	-0.0014	0.0095	0.900	0.593
6	100	0.2	0.1	10	GST	0.0256	0.0201	-0.0055	0.0068	0.914	0.548
6	100	0.2	0.1	10	WL	0.0256	0.0213	-0.0043	0.0058	0.922	0.565
7	40	0.2	0.2	5	HAD	0.0354	0.0354	-0.0000	0.0178	0.874	0.709
7	40	0.2	0.2	5	HC0	0.0354	0.0283	-0.0072	0.0147	0.857	0.639
7	40	0.2	0.2	5	HC1	0.0354	0.0323	-0.0031	0.0146	0.871	0.683
7	40	0.2	0.2	5	HC2	0.0354	0.0350	-0.0004	0.0164	0.875	0.707
7	40	0.2	0.2	5	HC3	0.0354	0.0438	0.0083	0.0251	0.890	0.785
7	40	0.2	0.2	5	MBN	0.0354	0.0366	0.0011	0.0146	0.887	0.730
7	40	0.2	0.2	5	GST	0.0354	0.0295	-0.0060	0.0103	0.917	0.663
7	40	0.2	0.2	5	WL	0.0354	0.0343	-0.0011	0.0092	0.937	0.715
8	70	0.2	0.2	5	HAD	0.0190	0.0191	0.0000	0.0056	0.923	0.532
8	70	0.2	0.2	5	HC0	0.0190	0.0169	-0.0021	0.0052	0.909	0.503
8	70	0.2	0.2	5	HC1	0.0190	0.0182	-0.0008	0.0051	0.919	0.522
8	70	0.2	0.2	5	HC2	0.0190	0.0189	-0.0001	0.0054	0.922	0.531

Scenario	n	Event proba-bility	Exposure prevalence	OR	Method	Mean true variance	Mean estimated variance	Bias	RMSE	95% CI coverage	Mean CI width
8	70	0.2	0.2	5	HC3	0.0190	0.0212	0.0022	0.0067	0.933	0.561
8	70	0.2	0.2	5	MBN	0.0190	0.0195	0.0004	0.0052	0.928	0.539
8	70	0.2	0.2	5	GST	0.0190	0.0158	-0.0032	0.0046	0.917	0.489
8	70	0.2	0.2	5	WL	0.0190	0.0172	-0.0019	0.0039	0.927	0.509
9	100	0.2	0.2	5	HAD	0.0131	0.0130	-0.0000	0.0029	0.931	0.443
9	100	0.2	0.2	5	HC0	0.0131	0.0120	-0.0010	0.0028	0.923	0.426
9	100	0.2	0.2	5	HC1	0.0131	0.0127	-0.0004	0.0027	0.929	0.437
9	100	0.2	0.2	5	HC2	0.0131	0.0130	-0.0001	0.0028	0.931	0.442
9	100	0.2	0.2	5	HC3	0.0131	0.0140	0.0010	0.0033	0.938	0.459
9	100	0.2	0.2	5	MBN	0.0131	0.0133	0.0002	0.0028	0.934	0.447
9	100	0.2	0.2	5	GST	0.0131	0.0108	-0.0023	0.0029	0.918	0.405
9	100	0.2	0.2	5	WL	0.0131	0.0114	-0.0016	0.0025	0.926	0.417
10	40	0.2	0.2	10	HAD	0.0358	0.0356	-0.0002	0.0176	0.869	0.710
10	40	0.2	0.2	10	HC0	0.0358	0.0286	-0.0072	0.0149	0.853	0.642
10	40	0.2	0.2	10	HC1	0.0358	0.0327	-0.0031	0.0148	0.865	0.686
10	40	0.2	0.2	10	HC2	0.0358	0.0352	-0.0005	0.0165	0.868	0.709
10	40	0.2	0.2	10	HC3	0.0358	0.0438	0.0080	0.0246	0.884	0.785
10	40	0.2	0.2	10	MBN	0.0358	0.0369	0.0012	0.0149	0.882	0.734
10	40	0.2	0.2	10	GST	0.0358	0.0298	-0.0059	0.0105	0.912	0.666
10	40	0.2	0.2	10	WL	0.0358	0.0347	-0.0011	0.0095	0.933	0.718
11	70	0.2	0.2	10	HAD	0.0193	0.0193	-0.0000	0.0059	0.917	0.535
11	70	0.2	0.2	10	HC0	0.0193	0.0172	-0.0021	0.0054	0.904	0.506
11	70	0.2	0.2	10	HC1	0.0193	0.0185	-0.0008	0.0054	0.913	0.525
11	70	0.2	0.2	10	HC2	0.0193	0.0191	-0.0002	0.0057	0.917	0.533
11	70	0.2	0.2	10	HC3	0.0193	0.0214	0.0021	0.0071	0.930	0.563
11	70	0.2	0.2	10	MBN	0.0193	0.0197	0.0004	0.0054	0.926	0.543
11	70	0.2	0.2	10	GST	0.0193	0.0161	-0.0032	0.0047	0.916	0.492
11	70	0.2	0.2	10	WL	0.0193	0.0174	-0.0019	0.0040	0.929	0.513
12	100	0.2	0.2	10	HAD	0.0131	0.0131	-0.0000	0.0028	0.932	0.444
12	100	0.2	0.2	10	HC0	0.0131	0.0121	-0.0010	0.0027	0.924	0.428
12	100	0.2	0.2	10	HC1	0.0131	0.0128	-0.0004	0.0027	0.930	0.439
12	100	0.2	0.2	10	HC2	0.0131	0.0130	-0.0001	0.0028	0.932	0.443
12	100	0.2	0.2	10	HC3	0.0131	0.0140	0.0009	0.0032	0.941	0.460
12	100	0.2	0.2	10	MBN	0.0131	0.0133	0.0002	0.0027	0.937	0.449
12	100	0.2	0.2	10	GST	0.0131	0.0109	-0.0022	0.0029	0.925	0.408
12	100	0.2	0.2	10	WL	0.0131	0.0116	-0.0015	0.0025	0.933	0.419
13	40	0.4	0.1	5	HAD	0.0709	0.0711	0.0002	0.0366	0.855	0.998
13	40	0.4	0.1	5	HC0	0.0709	0.0518	-0.0191	0.0303	0.839	0.865
13	40	0.4	0.1	5	HC1	0.0709	0.0592	-0.0117	0.0285	0.851	0.925
13	40	0.4	0.1	5	HC2	0.0709	0.0709	-0.0000	0.0331	0.856	1.004
13	40	0.4	0.1	5	HC3	0.0709	0.0988	0.0279	0.0603	0.870	1.172
13	40	0.4	0.1	5	MBN	0.0709	0.0690	-0.0020	0.0259	0.864	1.006
13	40	0.4	0.1	5	GST	0.0709	0.0682	-0.0027	0.0083	0.938	1.014
13	40	0.4	0.1	5	WL	0.0709	0.0788	0.0079	0.0126	0.953	1.089
14	70	0.4	0.1	5	HAD	0.0455	0.0452	-0.0003	0.0183	0.899	0.807
14	70	0.4	0.1	5	HC0	0.0455	0.0366	-0.0089	0.0167	0.880	0.733

Scenario	n	Event probability	Exposure prevalence	OR	Method	Mean true variance	Mean estimated variance	Bias	RMSE	95% CI coverage	Mean CI width
14	70	0.4	0.1	5	HC1	0.0455	0.0394	-0.0061	0.0160	0.891	0.761
14	70	0.4	0.1	5	HC2	0.0455	0.0452	-0.0003	0.0175	0.901	0.809
14	70	0.4	0.1	5	HC3	0.0455	0.0564	0.0109	0.0286	0.914	0.895
14	70	0.4	0.1	5	MBN	0.0455	0.0428	-0.0027	0.0147	0.902	0.795
14	70	0.4	0.1	5	GST	0.0455	0.0437	-0.0018	0.0039	0.939	0.808
14	70	0.4	0.1	5	WL	0.0455	0.0472	0.0017	0.0041	0.947	0.841
15	100	0.4	0.1	5	HAD	0.0314	0.0314	0.0000	0.0084	0.918	0.682
15	100	0.4	0.1	5	HC0	0.0314	0.0272	-0.0042	0.0082	0.903	0.637
15	100	0.4	0.1	5	HC1	0.0314	0.0286	-0.0028	0.0078	0.911	0.654
15	100	0.4	0.1	5	HC2	0.0314	0.0314	-0.0000	0.0082	0.918	0.682
15	100	0.4	0.1	5	HC3	0.0314	0.0364	0.0050	0.0130	0.932	0.731
15	100	0.4	0.1	5	MBN	0.0314	0.0302	-0.0012	0.0072	0.917	0.672
15	100	0.4	0.1	5	GST	0.0314	0.0302	-0.0012	0.0022	0.942	0.673
15	100	0.4	0.1	5	WL	0.0314	0.0319	0.0005	0.0019	0.948	0.692
16	40	0.4	0.1	10	HAD	0.0713	0.0712	-0.0002	0.0366	0.856	0.998
16	40	0.4	0.1	10	HC0	0.0713	0.0520	-0.0193	0.0305	0.837	0.867
16	40	0.4	0.1	10	HC1	0.0713	0.0594	-0.0119	0.0287	0.850	0.927
16	40	0.4	0.1	10	HC2	0.0713	0.0709	-0.0004	0.0332	0.858	1.004
16	40	0.4	0.1	10	HC3	0.0713	0.0985	0.0272	0.0600	0.871	1.170
16	40	0.4	0.1	10	MBN	0.0713	0.0692	-0.0021	0.0261	0.868	1.008
16	40	0.4	0.1	10	GST	0.0713	0.0688	-0.0025	0.0083	0.937	1.018
16	40	0.4	0.1	10	WL	0.0713	0.0795	0.0081	0.0129	0.953	1.093
17	70	0.4	0.1	10	HAD	0.0456	0.0457	0.0001	0.0174	0.905	0.813
17	70	0.4	0.1	10	HC0	0.0456	0.0371	-0.0086	0.0159	0.888	0.739
17	70	0.4	0.1	10	HC1	0.0456	0.0399	-0.0057	0.0151	0.898	0.767
17	70	0.4	0.1	10	HC2	0.0456	0.0456	0.0000	0.0166	0.907	0.814
17	70	0.4	0.1	10	HC3	0.0456	0.0568	0.0112	0.0278	0.922	0.900
17	70	0.4	0.1	10	MBN	0.0456	0.0433	-0.0023	0.0139	0.909	0.800
17	70	0.4	0.1	10	GST	0.0456	0.0438	-0.0018	0.0038	0.946	0.810
17	70	0.4	0.1	10	WL	0.0456	0.0474	0.0018	0.0041	0.953	0.842
18	100	0.4	0.1	10	HAD	0.0315	0.0315	-0.0001	0.0083	0.917	0.682
18	100	0.4	0.1	10	HC0	0.0315	0.0273	-0.0042	0.0082	0.903	0.638
18	100	0.4	0.1	10	HC1	0.0315	0.0287	-0.0028	0.0077	0.909	0.655
18	100	0.4	0.1	10	HC2	0.0315	0.0315	-0.0001	0.0081	0.918	0.682
18	100	0.4	0.1	10	HC3	0.0315	0.0364	0.0049	0.0127	0.931	0.731
18	100	0.4	0.1	10	MBN	0.0315	0.0303	-0.0012	0.0072	0.917	0.673
18	100	0.4	0.1	10	GST	0.0315	0.0303	-0.0012	0.0022	0.941	0.675
18	100	0.4	0.1	10	WL	0.0315	0.0320	0.0005	0.0019	0.947	0.693
19	40	0.4	0.2	5	HAD	0.0465	0.0468	0.0003	0.0159	0.916	0.830
19	40	0.4	0.2	5	HC0	0.0465	0.0378	-0.0086	0.0144	0.898	0.751
19	40	0.4	0.2	5	HC1	0.0465	0.0432	-0.0032	0.0129	0.914	0.803
19	40	0.4	0.2	5	HC2	0.0465	0.0466	0.0002	0.0144	0.919	0.830
19	40	0.4	0.2	5	HC3	0.0465	0.0580	0.0116	0.0250	0.934	0.921
19	40	0.4	0.2	5	MBN	0.0465	0.0497	0.0032	0.0128	0.931	0.862
19	40	0.4	0.2	5	GST	0.0465	0.0450	-0.0015	0.0055	0.941	0.823
19	40	0.4	0.2	5	WL	0.0465	0.0518	0.0053	0.0082	0.958	0.883

Scenario	n	Event proba-bility	Exposure prevalence	OR	Method	Mean true variance	Mean estimated variance	Bias	RMSE	95% CI coverage	Mean CI width
20	70	0.4	0.2	5	HAD	0.0248	0.0248	0.0000	0.0040	0.931	0.613
20	70	0.4	0.2	5	HC0	0.0248	0.0222	-0.0026	0.0043	0.918	0.581
20	70	0.4	0.2	5	HC1	0.0248	0.0239	-0.0009	0.0037	0.927	0.603
20	70	0.4	0.2	5	HC2	0.0248	0.0248	-0.0000	0.0038	0.931	0.613
20	70	0.4	0.2	5	HC3	0.0248	0.0277	0.0029	0.0055	0.943	0.646
20	70	0.4	0.2	5	MBN	0.0248	0.0258	0.0010	0.0037	0.938	0.626
20	70	0.4	0.2	5	GST	0.0248	0.0240	-0.0008	0.0020	0.941	0.604
20	70	0.4	0.2	5	WL	0.0248	0.0259	0.0011	0.0022	0.950	0.627
21	100	0.4	0.2	5	HAD	0.0169	0.0169	0.0000	0.0017	0.942	0.508
21	100	0.4	0.2	5	HC0	0.0169	0.0157	-0.0012	0.0019	0.934	0.490
21	100	0.4	0.2	5	HC1	0.0169	0.0166	-0.0004	0.0016	0.940	0.503
21	100	0.4	0.2	5	HC2	0.0169	0.0169	-0.0000	0.0016	0.942	0.508
21	100	0.4	0.2	5	HC3	0.0169	0.0182	0.0013	0.0022	0.950	0.527
21	100	0.4	0.2	5	MBN	0.0169	0.0174	0.0005	0.0017	0.945	0.516
21	100	0.4	0.2	5	GST	0.0169	0.0164	-0.0005	0.0011	0.946	0.500
21	100	0.4	0.2	5	WL	0.0169	0.0173	0.0003	0.0010	0.951	0.513
22	40	0.4	0.2	10	HAD	0.0466	0.0464	-0.0002	0.0159	0.906	0.827
22	40	0.4	0.2	10	HC0	0.0466	0.0378	-0.0088	0.0146	0.887	0.752
22	40	0.4	0.2	10	HC1	0.0466	0.0432	-0.0034	0.0130	0.904	0.804
22	40	0.4	0.2	10	HC2	0.0466	0.0463	-0.0003	0.0143	0.909	0.828
22	40	0.4	0.2	10	HC3	0.0466	0.0572	0.0106	0.0242	0.928	0.915
22	40	0.4	0.2	10	MBN	0.0466	0.0497	0.0031	0.0129	0.923	0.863
22	40	0.4	0.2	10	GST	0.0466	0.0450	-0.0016	0.0056	0.935	0.824
22	40	0.4	0.2	10	WL	0.0466	0.0518	0.0052	0.0082	0.949	0.883
23	70	0.4	0.2	10	HAD	0.0252	0.0252	-0.0000	0.0039	0.931	0.618
23	70	0.4	0.2	10	HC0	0.0252	0.0226	-0.0026	0.0043	0.919	0.586
23	70	0.4	0.2	10	HC1	0.0252	0.0244	-0.0008	0.0037	0.928	0.608
23	70	0.4	0.2	10	HC2	0.0252	0.0252	-0.0000	0.0038	0.931	0.617
23	70	0.4	0.2	10	HC3	0.0252	0.0281	0.0028	0.0055	0.944	0.651
23	70	0.4	0.2	10	MBN	0.0252	0.0263	0.0010	0.0037	0.939	0.631
23	70	0.4	0.2	10	GST	0.0252	0.0244	-0.0008	0.0020	0.939	0.609
23	70	0.4	0.2	10	WL	0.0252	0.0264	0.0011	0.0022	0.947	0.633
24	100	0.4	0.2	10	HAD	0.0171	0.0171	-0.0000	0.0017	0.939	0.511
24	100	0.4	0.2	10	HC0	0.0171	0.0160	-0.0012	0.0020	0.932	0.493
24	100	0.4	0.2	10	HC1	0.0171	0.0168	-0.0004	0.0016	0.938	0.506
24	100	0.4	0.2	10	HC2	0.0171	0.0171	-0.0000	0.0016	0.940	0.511
24	100	0.4	0.2	10	HC3	0.0171	0.0184	0.0013	0.0023	0.947	0.529
24	100	0.4	0.2	10	MBN	0.0171	0.0177	0.0005	0.0017	0.944	0.519
24	100	0.4	0.2	10	GST	0.0171	0.0166	-0.0005	0.0011	0.943	0.503
24	100	0.4	0.2	10	WL	0.0171	0.0175	0.0004	0.0010	0.949	0.517

Note. The exposure coefficient was set to its prespecified nonzero value. Bias is the mean estimated variance minus the mean analytic true variance. RMSE denotes root mean squared error. Coverage is evaluated among constructible two-sided 95% Wald confidence intervals using the standard normal critical value.

Supplementary Table S2. Type I error in simulations under the null hypothesis

Scenario	n	Event probability	Exposure prevalence	OR	Method	Type I error
1	40	0.2	0.1	5	HAD	0.248
1	40	0.2	0.1	5	HC0	0.238
1	40	0.2	0.1	5	HC1	0.202
1	40	0.2	0.1	5	HC2	0.191
1	40	0.2	0.1	5	HC3	0.145
1	40	0.2	0.1	5	MBN	0.086
1	40	0.2	0.1	5	GST	0.043
1	40	0.2	0.1	5	WL	0.031
2	70	0.2	0.1	5	HAD	0.234
2	70	0.2	0.1	5	HC0	0.241
2	70	0.2	0.1	5	HC1	0.232
2	70	0.2	0.1	5	HC2	0.224
2	70	0.2	0.1	5	HC3	0.209
2	70	0.2	0.1	5	MBN	0.185
2	70	0.2	0.1	5	GST	0.042
2	70	0.2	0.1	5	WL	0.034
3	100	0.2	0.1	5	HAD	0.160
3	100	0.2	0.1	5	HC0	0.165
3	100	0.2	0.1	5	HC1	0.161
3	100	0.2	0.1	5	HC2	0.158
3	100	0.2	0.1	5	HC3	0.153
3	100	0.2	0.1	5	MBN	0.154
3	100	0.2	0.1	5	GST	0.040
3	100	0.2	0.1	5	WL	0.036
4	40	0.2	0.1	10	HAD	0.237
4	40	0.2	0.1	10	HC0	0.229
4	40	0.2	0.1	10	HC1	0.190
4	40	0.2	0.1	10	HC2	0.177
4	40	0.2	0.1	10	HC3	0.132
4	40	0.2	0.1	10	MBN	0.079
4	40	0.2	0.1	10	GST	0.042
4	40	0.2	0.1	10	WL	0.031
5	70	0.2	0.1	10	HAD	0.240
5	70	0.2	0.1	10	HC0	0.242
5	70	0.2	0.1	10	HC1	0.234
5	70	0.2	0.1	10	HC2	0.228
5	70	0.2	0.1	10	HC3	0.215
5	70	0.2	0.1	10	MBN	0.187
5	70	0.2	0.1	10	GST	0.038
5	70	0.2	0.1	10	WL	0.031
6	100	0.2	0.1	10	HAD	0.162
6	100	0.2	0.1	10	HC0	0.168
6	100	0.2	0.1	10	HC1	0.164
6	100	0.2	0.1	10	HC2	0.160
6	100	0.2	0.1	10	HC3	0.154

Scenario	n	Event probability	Exposure prevalence	OR	Method	Type I error
6	100	0.2	0.1	10	MBN	0.156
6	100	0.2	0.1	10	GST	0.043
6	100	0.2	0.1	10	WL	0.037
7	40	0.2	0.2	5	HAD	0.149
7	40	0.2	0.2	5	HC0	0.155
7	40	0.2	0.2	5	HC1	0.131
7	40	0.2	0.2	5	HC2	0.129
7	40	0.2	0.2	5	HC3	0.103
7	40	0.2	0.2	5	MBN	0.075
7	40	0.2	0.2	5	GST	0.048
7	40	0.2	0.2	5	WL	0.035
8	70	0.2	0.2	5	HAD	0.103
8	70	0.2	0.2	5	HC0	0.110
8	70	0.2	0.2	5	HC1	0.102
8	70	0.2	0.2	5	HC2	0.100
8	70	0.2	0.2	5	HC3	0.092
8	70	0.2	0.2	5	MBN	0.089
8	70	0.2	0.2	5	GST	0.048
8	70	0.2	0.2	5	WL	0.041
9	100	0.2	0.2	5	HAD	0.073
9	100	0.2	0.2	5	HC0	0.081
9	100	0.2	0.2	5	HC1	0.073
9	100	0.2	0.2	5	HC2	0.072
9	100	0.2	0.2	5	HC3	0.065
9	100	0.2	0.2	5	MBN	0.066
9	100	0.2	0.2	5	GST	0.046
9	100	0.2	0.2	5	WL	0.040
10	40	0.2	0.2	10	HAD	0.146
10	40	0.2	0.2	10	HC0	0.150
10	40	0.2	0.2	10	HC1	0.126
10	40	0.2	0.2	10	HC2	0.124
10	40	0.2	0.2	10	HC3	0.100
10	40	0.2	0.2	10	MBN	0.075
10	40	0.2	0.2	10	GST	0.049
10	40	0.2	0.2	10	WL	0.036
11	70	0.2	0.2	10	HAD	0.097
11	70	0.2	0.2	10	HC0	0.105
11	70	0.2	0.2	10	HC1	0.097
11	70	0.2	0.2	10	HC2	0.094
11	70	0.2	0.2	10	HC3	0.086
11	70	0.2	0.2	10	MBN	0.084
11	70	0.2	0.2	10	GST	0.049
11	70	0.2	0.2	10	WL	0.041
12	100	0.2	0.2	10	HAD	0.070
12	100	0.2	0.2	10	HC0	0.078
12	100	0.2	0.2	10	HC1	0.070

Scenario	n	Event probability	Exposure prevalence	OR	Method	Type I error
12	100	0.2	0.2	10	HC2	0.068
12	100	0.2	0.2	10	HC3	0.061
12	100	0.2	0.2	10	MBN	0.061
12	100	0.2	0.2	10	GST	0.044
12	100	0.2	0.2	10	WL	0.039
13	40	0.4	0.1	5	HAD	0.170
13	40	0.4	0.1	5	HC0	0.187
13	40	0.4	0.1	5	HC1	0.174
13	40	0.4	0.1	5	HC2	0.167
13	40	0.4	0.1	5	HC3	0.148
13	40	0.4	0.1	5	MBN	0.147
13	40	0.4	0.1	5	GST	0.055
13	40	0.4	0.1	5	WL	0.043
14	70	0.4	0.1	5	HAD	0.109
14	70	0.4	0.1	5	HC0	0.125
14	70	0.4	0.1	5	HC1	0.116
14	70	0.4	0.1	5	HC2	0.107
14	70	0.4	0.1	5	HC3	0.097
14	70	0.4	0.1	5	MBN	0.106
14	70	0.4	0.1	5	GST	0.052
14	70	0.4	0.1	5	WL	0.045
15	100	0.4	0.1	5	HAD	0.083
15	100	0.4	0.1	5	HC0	0.095
15	100	0.4	0.1	5	HC1	0.090
15	100	0.4	0.1	5	HC2	0.082
15	100	0.4	0.1	5	HC3	0.070
15	100	0.4	0.1	5	MBN	0.082
15	100	0.4	0.1	5	GST	0.051
15	100	0.4	0.1	5	WL	0.045
16	40	0.4	0.1	10	HAD	0.157
16	40	0.4	0.1	10	HC0	0.174
16	40	0.4	0.1	10	HC1	0.161
16	40	0.4	0.1	10	HC2	0.153
16	40	0.4	0.1	10	HC3	0.137
16	40	0.4	0.1	10	MBN	0.129
16	40	0.4	0.1	10	GST	0.050
16	40	0.4	0.1	10	WL	0.036
17	70	0.4	0.1	10	HAD	0.113
17	70	0.4	0.1	10	HC0	0.130
17	70	0.4	0.1	10	HC1	0.121
17	70	0.4	0.1	10	HC2	0.111
17	70	0.4	0.1	10	HC3	0.097
17	70	0.4	0.1	10	MBN	0.110
17	70	0.4	0.1	10	GST	0.051
17	70	0.4	0.1	10	WL	0.044
18	100	0.4	0.1	10	HAD	0.086

Scenario	n	Event probability	Exposure prevalence	OR	Method	Type I error
18	100	0.4	0.1	10	HC0	0.099
18	100	0.4	0.1	10	HC1	0.092
18	100	0.4	0.1	10	HC2	0.085
18	100	0.4	0.1	10	HC3	0.074
18	100	0.4	0.1	10	MBN	0.085
18	100	0.4	0.1	10	GST	0.050
18	100	0.4	0.1	10	WL	0.046
19	40	0.4	0.2	5	HAD	0.095
19	40	0.4	0.2	5	HC0	0.114
19	40	0.4	0.2	5	HC1	0.096
19	40	0.4	0.2	5	HC2	0.092
19	40	0.4	0.2	5	HC3	0.074
19	40	0.4	0.2	5	MBN	0.077
19	40	0.4	0.2	5	GST	0.057
19	40	0.4	0.2	5	WL	0.044
20	70	0.4	0.2	5	HAD	0.070
20	70	0.4	0.2	5	HC0	0.081
20	70	0.4	0.2	5	HC1	0.073
20	70	0.4	0.2	5	HC2	0.069
20	70	0.4	0.2	5	HC3	0.058
20	70	0.4	0.2	5	MBN	0.063
20	70	0.4	0.2	5	GST	0.053
20	70	0.4	0.2	5	WL	0.046
21	100	0.4	0.2	5	HAD	0.064
21	100	0.4	0.2	5	HC0	0.073
21	100	0.4	0.2	5	HC1	0.066
21	100	0.4	0.2	5	HC2	0.064
21	100	0.4	0.2	5	HC3	0.055
21	100	0.4	0.2	5	MBN	0.059
21	100	0.4	0.2	5	GST	0.053
21	100	0.4	0.2	5	WL	0.047
22	40	0.4	0.2	10	HAD	0.093
22	40	0.4	0.2	10	HC0	0.113
22	40	0.4	0.2	10	HC1	0.094
22	40	0.4	0.2	10	HC2	0.089
22	40	0.4	0.2	10	HC3	0.073
22	40	0.4	0.2	10	MBN	0.075
22	40	0.4	0.2	10	GST	0.056
22	40	0.4	0.2	10	WL	0.041
23	70	0.4	0.2	10	HAD	0.072
23	70	0.4	0.2	10	HC0	0.084
23	70	0.4	0.2	10	HC1	0.074
23	70	0.4	0.2	10	HC2	0.071
23	70	0.4	0.2	10	HC3	0.061
23	70	0.4	0.2	10	MBN	0.066
23	70	0.4	0.2	10	GST	0.056

Scenario	n	Event probability	Exposure prevalence	OR	Method	Type I error
23	70	0.4	0.2	10	WL	0.049
24	100	0.4	0.2	10	HAD	0.063
24	100	0.4	0.2	10	HC0	0.070
24	100	0.4	0.2	10	HC1	0.064
24	100	0.4	0.2	10	HC2	0.063
24	100	0.4	0.2	10	HC3	0.056
24	100	0.4	0.2	10	MBN	0.059
24	100	0.4	0.2	10	GST	0.052
24	100	0.4	0.2	10	WL	0.046

Note. The exposure coefficient was set to zero. The intercept was recalibrated to preserve the target marginal event probability. Type I error is based on the two-sided Wald test using the standard normal critical value.

Supplementary Table S3. Computational feasibility and diagnostic outcomes

Panel A. Computational feasibility and causes of non-computability

n	Variance estimator	Computable analyses (%)	Rank-deficient design (%)	Hadamard-conditioning failure (%)	Leverage-related failure (%)
40	HAD	88.2	0.75	11.01	0.00
40	HC0	99.3	0.75	0.00	0.00
40	HC1	99.3	0.75	0.00	0.00
40	HC2	96.9	0.75	0.00	2.36
40	HC3	96.9	0.75	0.00	2.36
40	MBN	99.3	0.75	0.00	0.00
40	GST	99.3	0.75	0.00	0.00
40	WL	96.9	0.75	0.00	2.37
70	HAD	98.8	0.02	1.16	0.00
70	HC0	100.0	0.02	0.00	0.00
70	HC1	100.0	0.02	0.00	0.00
70	HC2	99.9	0.02	0.00	0.12
70	HC3	99.9	0.02	0.00	0.12
70	MBN	100.0	0.02	0.00	0.00
70	GST	100.0	0.02	0.00	0.00
70	WL	99.9	0.02	0.00	0.12
100	HAD	99.9	0.00	0.09	0.00
100	HC0	100.0	0.00	0.00	0.00
100	HC1	100.0	0.00	0.00	0.00
100	HC2	100.0	0.00	0.00	0.01
100	HC3	100.0	0.00	0.00	0.01
100	MBN	100.0	0.00	0.00	0.00
100	GST	100.0	0.00	0.00	0.00
100	WL	100.0	0.00	0.00	0.01

Note. Results are pooled across simulations with a nonzero exposure effect and simulations under the null hypothesis. Percentages use all attempted datasets as the denominator. Hadamard-conditioning failure denotes designs for which the estimated reciprocal 1-norm condition number of the Hadamard matrix $\boldsymbol{A}$, computed using the R function `rcond()`, was below 10^{-12} or could not be computed. Leverage-related failure denotes

method-specific designs for which the required leverage adjustment was not numerically evaluable.

Panel B. Diagnostic outcomes for the HAD estimator

n	Any negative component (%)	Any component > 0.25 (%)	Non-PSD covariance (%)	Nonpositive target variance (%)
40	85.2	100.0	15.3	0.202
70	58.8	100.0	5.1	0.018
100	38.4	100.0	2.1	0.002

Note. Results are pooled across simulations with a nonzero exposure effect and simulations under the null hypothesis. Percentages are calculated among computable HAD analyses. Negative variance components, components exceeding 0.25, non-positive-semidefinite covariance matrices, and nonpositive target variances were recorded as diagnostic outcomes rather than computational failures.

Supplementary Table S4. Sensitivity analysis of stabilized HAD estimators

n	Method	Mean true variance, range	Mean estimated variance, range	Bias, range	RMSE, range
40	HAD	0.0354–0.0713	0.0354–0.0712	-0.0002 to 0.0003	0.0159–0.0383
40	HAD0	0.0354–0.0713	0.0365–0.0759	0.0008 to 0.0050	0.0160–0.0405
40	HAD025	0.0354–0.0713	0.0189–0.0362	-0.0351 to -0.0166	0.0207–0.0413
70	HAD	0.0190–0.0456	0.0191–0.0457	-0.0003 to 0.0001	0.0039–0.0205
70	HAD0	0.0190–0.0456	0.0192–0.0466	0.0000 to 0.0010	0.0039–0.0210
70	HAD025	0.0190–0.0456	0.0111–0.0275	-0.0184 to -0.0071	0.0084–0.0244
100	HAD	0.0131–0.0315	0.0130–0.0315	-0.0001 to 0.0000	0.0017–0.0110
100	HAD0	0.0131–0.0315	0.0131–0.0317	0.0000 to 0.0003	0.0017–0.0110
100	HAD025	0.0131–0.0315	0.0078–0.0213	-0.0109 to -0.0041	0.0047–0.0149

Note. Results are from simulations with a nonzero exposure effect. HAD denotes the untruncated estimator. HAD0 replaces negative estimated variance components by zero, whereas HAD025 truncates estimated variance components to the interval [0, 0.25]. Ranges are calculated across the eight simulation scenarios within each sample size.