Biometrics 71, 704–713
September 2015

DOI: 10.1111/biom.12318

The Direct Integral Method for Confidence Intervals for the Ratio of Two Location Parameters

Yanqing Wang,[1] Suojin Wang,[2] and Raymond J. Carroll[2,3]*

[1]Department of Biostatistics, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, U.S.A.
[2]Department of Statistics, Texas A&M University, 3143 TAMU, College Station, Texas 77843-3143, U.S.A.
[3]Department of Mathematics and Statistics, University of Technology, Sydney, Australia
**email:* carroll@stat.tamu.edu

Summary. In a relative risk analysis of colorectal caner on nutrition intake scores across genders, we show that, surprisingly, when comparing the relative risks for men and women based on the index of a weighted sum of various nutrition scores, the problem reduces to forming a confidence interval for the ratio of two (asymptotically) normal random variables. The latter is an old problem, with a substantial literature. However, our simulation results suggest that existing methods often either give inaccurate coverage probabilities or have a positive probability to produce confidence intervals with infinite length. Motivated by such a problem, we develop a new methodology which we call the Direct Integral Method for Ratios (DIMER), which, unlike the other methods, is based directly on the distribution of the ratio. In simulations, we compare this method to many others. These simulations show that, generally, DIMER more closely achieves the nominal confidence level, and in those cases that the other methods achieve the nominal levels, DIMER has comparable confidence interval lengths. The methodology is then applied to a real data set, and with follow up simulations.

Key words: Confidence interval; Direct Integral Method for Ratios; DIMER; Fieller's interval; Hayya's method; Ratios of location parameters.

1. Introduction

We use data on the relationship between diet and colorectal cancer from a subset of the NIH-AARP Study of Diet and Health (Reedy et al., 2008), which itself is a large cohort study with approximately 250,000 men and 200,000 women. The data subset that we have access to includes 1075 males that developed colorectal cancer during the course of the study, along with 479 females who also developed colorectal cancer. In addition, the data set includes 3225 randomly selected men and 1437 randomly selected women who did not develop colorectal cancer. Hence, there are 4300 males and 1916 females in the data set.

It is traditional in nutritional epidemiology to examine the risk of cancer from single foods or nutrients normalized by energy (caloric) intake, for example, the percentage of calories coming from fat, the amount of whole grants per 1000 calories, etc. However, nutritionists have increasingly turned to dietary *indices*, which account for the patterns of energy-adjusted intake for multiple foods and nutrients. There are many such indices, for example, the Healthy Eating Index-2005 (HEI-2005, see Guenther, Reedy, and Krebs-Smith, 2008), the Alternative Healthy Eating Index, the Mediterranean Index, etc., and they have been shown to be related to many chronic diseases and cancers. We use here the HEI-2005, which is based on the intakes of 12 interrelated dietary components, adjusted for energy intake. These intakes are then scored individually, and their sum is the HEI-2005, which is then used to predict disease. The **Supplementary Material** Table S.3 describes the components and how they are scored.

In our analysis of colorectal cancer, we fit a model where the scores are weighted and summed, but the weights are common for men and women, as in any dietary index. We show in Section 4.1 that, surprisingly, when comparing the relative risks for men and women based on this common index, the problem reduces to forming a confidence interval for the ratio of two (asymptotically) normal random variables. The latter is an old problem, with a substantial literature, one that we revisit based on our example.

One popular method for computing a confidence interval for the ratio of two location parameters is due to Fieller (1932, 1954). Details of this method are described in the **Supplementary Material** Appendix S.1.

Consequently, other methods have been developed, most of which are based on the distribution of the ratio of the estimates of two location parameters (see, e.g., recent articles by Beyene and Moineddin, 2005; Pham-Gia, Turkkan, and Marchand, 2006; Sherman, Maity, and Wang, 2011). Most often, a normal approximation to the distribution is used, with subsequent intervals formed by Wald's method. Hayya, Armstrong, and Gressis (1975) showed that, under certain conditions, the distribution for the ratio of two estimators can be treated as a normal distribution with a second order Taylor expansion. This method is also defined in the **Supplementary Material** Appendix S.3. In addition, parametric and nonparametric bootstrap methods are also used. However, our empirical investigations suggest that confidence intervals constructed by these existing methods for the ratio often give inaccurate and sub-nominal coverage probabilities.

Motivated by such a problem, in this work we construct a new methodology, which we call the *Direct Integral Method for Ratios (DIMER)*. This methodology is also based on the distribution of the ratio of the estimates of the two location parameters, a distribution that is Cauchy-like and has heavy tails. We show that DIMER can be computed easily by numerical integration. In our simulation studies, we show that DIMER closely achieves nominal coverage, unlike the Wald method and the method of Hayya et al. (1975). DIMER is also much faster computationally than bootstrap methods, which is important in large cohort studies, where the model is a nonlinear logistic regression based on samples of sizes in the tens of thousands or more.

In Section 2, we describe the methodology, while Section 3 compares various methods via simulation studies. Section 4 describes the analysis of the NIH-AARP Study. Simulations based on the actual data reinforce the conclusions of the simulations in Section 3 and shed more light on the data analysis. Technical details, proofs, definitions and additional simulations are given in the **Supplementary Material**.

2. Methodology

2.1. *Basic Definitions*

Consider two random variables T_1 and T_2 which have density functions $f_1\{(t_1 - \mu_1)/v_1\}$ and $f_2\{(t_2 - \mu_2)/v_2\}$, respectively, with means μ_1 and μ_2 and standard deviations v_1 and v_2. In other words, $f_1(x)$ and $f_2(x)$ are the density functions of the standardized versions of T_1 and T_2, respectively. Let $F_1(\cdot)$ and $F_2(\cdot)$ denote the corresponding distribution functions. We are interested in making inference for the ratio μ_1/μ_2. We will outline a series of cases where it is possible to compute easily the cumulative distribution function of $\hat{r} = T_1/T_2$. All proofs are given in the **Supplementary Material** Section S.2.

2.2. *Independent Case*

Suppose that T_1 and T_2 are independent.

LEMMA 1. *Define*

$$g(z|x, \mu_1, \mu_2, v_1, v_2) = \begin{cases} (1 - F_1[\{x(\mu_2 + v_2 z) - \mu_1\}/v_1]) f_2(z)\exp(z^2) & \text{if } z \le -\mu_2/v_2, \\ F_1[\{x(\mu_2 + v_2 z) - \mu_1\}/v_1] f_2(z)\exp(z^2) & \text{if } z > -\mu_2/v_2. \end{cases}$$

Then the cumulative distribution function of $\hat{r} = T_1/T_2$ is given by

$$pr(\hat{r} \le x) = \int_{-\infty}^{\infty} g(z|x, \mu_1, \mu_2, v_1, v_2)\exp(-z^2)\mathrm{d}z,$$

a quantity that is easily computed by Gauss–Hermite quadrature.

In Lemma 1 x denotes a value of $\hat{r}$ and z denotes a value of T_2, and similarly in Sections 2.3–2.4.

If the parameters v_1 and v_2 are unknown, we can apply Lemma 1 using their estimated values. However, we have found that a more numerically efficient approximation can be developed in the case of normally distributed T_1 and T_2. We present this result in the following setting. Suppose the estimated variances are $\hat{v}_1^2$ and $\hat{v}_2^2$ which are independent of each other, and independent of T_1 and T_2, and have degrees of freedom d_1 and d_2, respectively. Thus, both $(T_1 - \mu_1)/\hat{v}_1$ and $(T_2 - \mu_2)/\hat{v}_2$ follow the t-distribution with d_1 and d_2 degrees of freedom, respectively. In addition, assume that $d = \min(d_1, d_2)$ increases to infinity, which is implied when the sample sizes increase to infinity. Suppose that $\hat{v}_1^2 = v_1^2 + O_p(d_1^{-1/2})$ and $\hat{v}_2^2 = v_2^2 + O_p(d_2^{-1/2})$. Then we have the following lemma.

LEMMA 2. *With an error of order $O_p(d^{-1/2})$, $g(z|x, \mu_1, \mu_2, \hat{v}_1^2, \hat{v}_2^2)$ defined in Lemma 1 can be approximated by*

$$h(z|x, \mu_1, \mu_2, \hat{v}_1^2, \hat{v}_2^2) = \begin{cases} (1 - F_{t,d_1}[\{x(\mu_2 + \hat{v}_2 z) - \mu_1\}/\hat{v}_1]) f_{t,d_2}(z)\exp(z^2) & \text{if } z \le -\mu_2/\hat{v}_2, \\ F_{t,d_1}[\{x(\mu_2 + \hat{v}_2 z) - \mu_1\}/\hat{v}_1] f_{t,d_2}(z)\exp(z^2) & \text{if } z > -\mu_2/\hat{v}_2, \end{cases}$$

where $f_{t,d}(\cdot)$ and $F_{t,d}(\cdot)$ are the t-density with d degrees of freedom and the corresponding cumulative distribution function, respectively.

2.3. *Dependent Case of Two Normally Distributed Variables with Known Covariance Matrix*

Suppose now that (T_1, T_2) are jointly normally distributed with means (μ_1, μ_2), variances (v_1^2, v_2^2), covariance v_{12} and suppose that (v_1^2, v_2^2, v_{12}) are known. Let $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal density and distribution function.

LEMMA 3. *Define $g(z|x, \mu_1, \mu_2, v_1^2, v_2^2, v_{12})$ as follows. If $z \le -\mu_2/v_2$, then*

$$g(z|x, \mu_1, \mu_2, v_1^2, v_2^2, v_{12}) = (2\pi)^{-1/2}(1 - \Phi[\{x(\mu_2 + v_2 z) - (\mu_1 + z v_{12}/v_2)\}v_2/\sqrt{v_1^2 v_2^2 - v_{12}^2}])\exp(z^2/2).$$

If $z > -\mu_2/v_2$, then

$$g(z|x, \mu_1, \mu_2, v_1^2, v_2^2, v_{12}) = (2\pi)^{-1/2}\Phi[\{x(\mu_2 + v_2 z) - (\mu_1 + z v_{12}/v_2)\}v_2/\sqrt{v_1^2 v_2^2 - v_{12}^2}]\exp(z^2/2).$$

Then the distribution function of $\hat{r}$ is

$$pr(\hat{r} \le x) = \int_{-\infty}^{\infty} g(z|x, \mu_1, \mu_2, v_1^2, v_2^2, v_{12})\exp(-z^2)\mathrm{d}z,$$

which again can be computed by Gauss–Hermite quadrature.

Of course, when $v_{12} = 0$, Lemma 3 is a special case of Lemma 1.

2.4. *Dependent Case of Two Normally Distributed Variables with Estimated Covariance Matrix*

Here, we discuss the cumulative distribution of the ratio $\hat{r} = T_1/T_2$ when T_1 and T_2 are jointly normally distributed with jointly estimated variance and covariance which have the same number of degrees of freedom d, and these estimates are independent of T_1 and T_2. These are the same assumptions noted in Fieller (1954). Define the estimates of the variances and covariance of T_1 and T_2 as $\hat{v}_1^2$, $\hat{v}_2^2$, and $\hat{v}_{12}$. Let $\eta = v_{12}/v_2^2$. For fixed η, write $W = T_1 - \eta T_2$, Then W and T_2 are independent. In addition, if $\hat{v}_1^2$, $\hat{v}_2^2$ and $\hat{v}_{12}$ are computed from the sample covariance matrix of normal random variables from a sample of size $d+1$, then we also have that $T_1 - \eta T_2$ and T_2 are independent of their estimated variances $\hat{v}_1^2 - 2\eta\hat{v}_{12} + \eta^2\hat{v}_2^2$ and $\hat{v}^2$, which are independent of each other and also have d degrees of freedom.

We use the following algorithm, based on the approximation used in Section 2.2. Under our assumptions, the variables $Z_1 = \{(T_1 - \eta T_2) - (\mu_1 - \eta\mu_2)\}/\sqrt{\hat{v}_1^2 - 2\eta\hat{v}_{12} + \eta^2\hat{v}_2^2}$ and $Z_2 = (T_2 - \mu_2)/\hat{v}_2$ are independent and both have t-distributions with d degrees of freedom. As in Lemma 2, we then make the approximation that the density of (T_1, T_2), having fixed the estimated covariance matrix, is approximately

$$\hat{v}_2^{-1}(\hat{v}_1^2 - 2\eta\hat{v}_{12} + \eta^2\hat{v}_2^2)^{-1/2} f_{t,d}[\{(t_1 - \eta t_2) \\ -(\mu_1 - \eta\mu_2)\}/\sqrt{\hat{v}_1^2 - 2\eta\hat{v}_{12} + \eta^2\hat{v}_2^2}] f_{t,d}\{(t_2 - \mu_2)/\hat{v}_2\}.$$

If $z \le -\mu_2/\hat{v}_2$, define

$$\begin{aligned} &g(z|x, \mu_1, \mu_2, \hat{v}_1^2, \hat{v}_2^2, \hat{v}_{12}, \eta) \\ &\quad = \left(1 - F_{t,d}\left[\{(x-\eta)(\mu_2 + v_2 z) \right.\right. \\ &\qquad \left.\left. -(\mu_1 - \eta\mu_2)\}/\sqrt{\hat{v}_1^2 - 2\eta\hat{v}_{12} + \eta^2\hat{v}_2^2}\right]\right) f_{t,d}(z)\exp(z^2), \end{aligned}$$

while if $z > -\mu_2/\hat{v}_2$, define

$$\begin{aligned} &g(z|x, \mu_1, \mu_2, \hat{v}_1^2, \hat{v}_2^2, \hat{v}_{12}, \eta) \\ &\quad = F_{t,d}\left[\{(x-\eta)(\mu_2 + v_2 z) \right. \\ &\qquad \left. -(\mu_1 - \eta\mu_2)\}/\sqrt{\hat{v}_1^2 - 2\eta\hat{v}_{12} + \eta^2\hat{v}_2^2}\right] f_{t,d}(z)\exp(z^2). \end{aligned}$$

Then, using the same device as in Lemma 2, we have that

$$\begin{aligned} \mathrm{pr}(\hat{r} \le x) = &\int_{-\infty}^{\infty} g(z|x, \mu_1, \mu_2, \hat{v}_1^2, \hat{v}_2^2, \hat{v}_{12}, \eta)\exp(-z^2)\mathrm{d}z \\ &+ O_p(d^{-1/2}). \end{aligned} \tag{1}$$

In practice, η is unknown, so we use $\hat{\eta} = \hat{v}_{12}/\hat{v}_2^2$ to estimate it.

2.5. *Algorithm for Computing the Confidence Interval of Ratios*

In Sections 2.2–2.4, we express the distribution function of $\hat{r}$ as $F(x; r) = \mathrm{pr}(\hat{r} \le x; r = \mu_1/\mu_2)$ when $\mu_2 \ne 0$. The ratio $\hat{\mu}_1/\hat{\mu}_2$ is an estimate of $r = \mu_1/\mu_2$, so that we can view $F(x; \hat{\mu}_1/\hat{\mu}_2)$ as an estimate of the population distribution function $F(x; r)$. Efron (1981) and Benton and Krishnamoorthy (2002) pointed out that if we generate values $\hat{r}_i$, $i = 1, \ldots, m$, from $F(x; \hat{\mu}_1/\hat{\mu}_2)$, we can make inference about r using the distribution of the generated $\hat{r}_i$'s.

The main difference between our approach and that of Benton and Krishnamoorthy is that instead of generating a larger number of $\hat{r}_i$'s and then obtaining its percentiles, we compute the percentile of $\hat{r}_i$ directly. Consequently, our method is much faster computationally. Specifically, our simulation results indicate that DIMER usually needs less than 30 iteration steps to obtain the quantile of a distribution, but in Benton and Krishnamoorthy (2002), they used $m = 100{,}000$ $\hat{r}_i$'s to get the quantiles.

Define the $\alpha/2$ quantile for $F(x; \hat{\mu}_1/\hat{\mu}_2)$ as $\hat{r}_{\alpha/2|\hat{\mu}_1/\hat{\mu}_2}$. Then an approximate $100(1-\alpha)\%$ confidence interval for r is $(\hat{r}_{\alpha/2|\hat{\mu}_1/\hat{\mu}_2}, \hat{r}_{1-\alpha/2|\hat{\mu}_1/\hat{\mu}_2})$. Here, we give the steps of our iterative, bisection-based algorithm to obtain the quantiles.

- Step 1. Give two initial values of $\hat{r}_{\alpha/2|\hat{\mu}_1/\hat{\mu}_2}$ as $\hat{r}_{\alpha_1} < 0 < \hat{r}_{\alpha_2}$ and both have sufficiently large absolute values to make sure that $\hat{r}_{\alpha/2|\hat{\mu}_1/\hat{\mu}_2}$ is inside the interval $(\hat{r}_{\alpha_1}, \hat{r}_{\alpha_2})$. How we did this is described in the **Supplementary Material** Appendix S.4. Our method, being based on bisection, is not sensitive to these starting values.
- Step 2. Apply the Gauss–Hermit quadrature to the cumulative distribution function of $\hat{r}$ to obtain $c_{\alpha/2} = \mathrm{pr}\{\hat{r} \le (\hat{r}_{\alpha_1} + \hat{r}_{\alpha_2})/2\}$. If $c_{\alpha/2} < \alpha/2$, let $\hat{r}_{\alpha_1} = (\hat{r}_{\alpha_1} + \hat{r}_{\alpha_2})/2$; if $c_{\alpha/2} > \alpha/2$, let $\hat{r}_{\alpha_2} = (\hat{r}_{\alpha_1} + \hat{r}_{\alpha_2})/2$; if $c_{\alpha/2} = \alpha/2$, stop the iteration and let $\hat{r}_{\alpha/2|\hat{\mu}_1/\hat{\mu}_2} = (\hat{r}_{\alpha_1} + \hat{r}_{\alpha_2})/2$.
- Step 3. Repeat Step 2 until $c_{\alpha/2}$ is close to $\alpha/2$ and/or the difference $|\hat{r}_{\alpha_2} - \hat{r}_{\alpha_1}|$ is negligible. Then we have $\hat{r}_{\alpha/2|\hat{\mu}_1/\hat{\mu}_2} = (\hat{r}_{\alpha_1} + \hat{r}_{\alpha_2})/2$, the lower limit of interval.
- Step 4. Repeat Steps 1–3 to obtain $\hat{r}_{1-\alpha/2|\hat{\mu}_1/\hat{\mu}_2}$, the upper limit of the interval.

3. Simulations

3.1. *Overview*

We performed simulations on two simple linear regression models. The first (Section 3.4) is to illustrate an application of the formulas in Section 2.2 where the two variables are independent. The second (in **Supplementary Material** Appendix Section S.6) is an example to demonstrate the performance of our method developed in Section 2.4 when the two variables are dependent. In both simulations, some other possible methods are outlined and compared with DIMER. Since the dependent case is developed with the normality assumption, it is important to evaluate how sensitive DIMER is to the violation of this assumption. Therefore, we also considered such a case in the second part of our simulations.

3.2. *Comments Upon and Applications of Fieller's Intervals*

Fieller's interval, defined in **Supplementary Material** Section S.1, is sometimes of infinite length, being either the entire real line or the union of two disconnected infinite length intervals, for example, when the denominator of the ratio is not significantly different from zero.

Fieller's intervals have been used in a variety of contexts. Here are three cases, the first two of which are illustrated in our simulations. The simulation of the first case appears here, while the second in the **Supplementary Material**, Section S.6.

- In a slope ratio assay (Finney, 1978; Hubert, 1984; Redmond, 2005c), data are fit to a standard and treatment, observing $Y_S = \alpha_S + X_S\beta_S + \epsilon_S$ for the standard, while the treatment is fit to the model $Y_T = \alpha_T + X_T\beta_T + \epsilon_T$. The relative potency ρ is a function of β_T/β_S, where the estimates of β_T and β_S are independent. In a common setting, it is assumed that $\alpha_S = \alpha_T$ but the doses $X_S = X_T = X$, and by centering X the described model holds with different intercepts. This is an example of two independent slope estimates.
- In a radioimmunoassay (Finney, 1978; Redmond, 2005b) with dose denoted by X and response Y, if one is in the linear part of the calibration cure a reasonable model is $Y = \alpha + X\beta + \epsilon$. The logarithm of ID_{50}, the dose required for 50% of binding inhibition, is given by $\log(\mathrm{ID}_{50}) = \alpha/\beta$. The parameter estimates $(\hat{\alpha}, \hat{\beta})$ are generally correlated, and this is an example of estimating the ratio of the intercept to the slope when the parameter estimates are correlated.
- In a parallel line assay (Finney, 1978; Redmond, 2005a), a standard is fit to the linear model $Y_S = \alpha_S + X_S\beta + \epsilon_S$ while the treatment is fit to the model $Y_T = \alpha_T + X_T\beta + \epsilon_T$: the slope is the same in both, hence parallel line. The log-relative potency in this assay is $\log(\rho) = (\alpha_T - \alpha_S)/\beta$. In the homoscedastic case, unless $X_S = X_T$, the estimates of the numerator and denominator are not independent.

In radioimmunoassays, it is often the case that the variance of the responses is proportional to a power θ of the mean, but with $1 < \theta < 2$. Generalized least squares can then be used to estimate θ (Davidian, Carroll, and Smith, 1988), but once the estimates in these examples are obtained, we still have a problem of forming a confidence interval for a ratio of two parameters.

3.3. *Comments on Sample Sizes and Parameter Choices*

Fieller intervals for a ratio θ_1/θ_2 are of infinite length if the null hypothesis $H_0 : \theta_2 = 0$ cannot be rejected. If the power for rejecting this hypothesis is low, Fieller intervals will have terrible properties. In simulations not reported here, the behavior of the alternative methods is also very poor. If the power for rejecting the hypothesis is essentially 100%, then all the methods will be essentially the same, with minor fluctuations depending on the sample size. The interesting cases lie on the boundary between low and perfect power, for example, 80–90% power with Type I error 0.05. Our simulations include settings with low power, perfect power and in between.

In our simulations, which are based on linear regression with error standard deviation v_ϵ, we have set the covariates to be Normal(0, 1), and we set $v_\epsilon = 1$, so that the standard error of the slope is roughly $n^{-1/2}v_\epsilon/s_x$, where s_x is the sample standard deviation of the covariates. On average , $s_x^{-1} \approx 1.0$, so the standard error of the slope estimate $\approx n^{-1/2}$. Consequently, the sample sizes we have chosen, n = 18, 25, and 50, result in reasonable standard errors that illustrate a range of powers for the test that the slope = 0.0. In Table 2, had we changed $v_\epsilon = 2$, 3, and 4, the sample sizes needed to get roughly the same percentage of infinite length Fieller intervals are roughly 60, 130, and 225, respectively. In the **Supplementary Material**, Table S.4, we show what happens to Table 1 when we set $(n, v_\epsilon) = (55, 2)$ and $(115, 3)$, showing that roughly the same results apply in this setting.

3.4. *Linear Model When the Two Estimates are Independent*

3.4.1. Setup. Consider the 2-group linear regression model

$$
\begin{aligned}
Y_{1i} &= \beta_{10} + X_{1i}\beta_{11} + \epsilon_{1i},\ i = 1, \ldots, n_1;\\
Y_{2j} &= \beta_{20} + X_{2j}\beta_{21} + \epsilon_{2j},\ j = 1, \ldots, n_2,
\end{aligned}
$$

where (Y_{1i}, X_{1i}) and (Y_{2j}, X_{2j}) are the same outcomes and predictors from different populations. See Section 3.2 for an example. Also ϵ_{1i} and ϵ_{2j} are independently normally distributed with mean zero and variances $v_{\epsilon_1}^2$ and $v_{\epsilon_2}^2$, respectively. Our interest is in the ratio of the two slopes β_{21}/β_{11}.

The model can be rewritten as follows in order to use a simple expression for the ratio:

$$
\begin{aligned}
Y_{1i} &= \beta_{10} + X_{1i}\omega + \epsilon_{1i},\ i = 1, \ldots, n_1;\\
Y_{2j} &= \beta_{20} + \beta_{21}X_{2j}\omega + \epsilon_{2j},\ j = 1, \ldots, n_2.
\end{aligned}
\tag{2}
$$

Then the ratio of the slopes now is β_{21} and ω is the slope for the first group..

Our interest is to construct a confidence interval for β_{21}. Let $(\hat{\beta}_{21}, \hat{\omega})$ denote the maximum likelihood estimate (mle) of (β_{21}, ω), and define $\lambda = \beta_{21}\omega$ and its estimate $\hat{\lambda} = \hat{\beta}_{21}\hat{\omega}$. Both $(\hat{\lambda} - \lambda)/\hat{v}_\lambda$ and $(\hat{\omega} - \omega)/\hat{v}_\omega$ follow independent t-distributions with degrees of freedom $n_2 - 2$ and $n_1 - 2$, respectively, where $\hat{v}_\lambda$ and $\hat{v}_\omega$ are corresponding estimated standard deviations.

The estimated cumulative distribution function of $\hat{\beta}_{21}$ is obtained as in Section 2.2. We can then apply the DIMER algorithm in Section 2.5 to obtain confidence intervals. To compare with other methods, in Section 3.4.2 we outline an application of Fieller's interval. In addition, we apply the Wald interval by inverting the Fisher score matrix, Hayya's method, the nonparametric bootstrap, the parametric bootstrap, and the likelihood ratio test; see the details in the **Supplementary Material**, Appendices S.1, S.3, and S.5.

3.4.2. Comparison with the Fieller's Interval. To form a confidence interval for β_{21}, one common method in practice is Fieller's interval. However, in this linear regression setting, it cannot be applied directly since $\hat{v}_{\hat{\omega}}^2$ and $\hat{v}_{\hat{\lambda}}^2$ are obtained independently. In this case, by the Welch–Satterthwaite equation (Satterthwaite, 1946; Welch, 1947), the degrees of freedom of $(\hat{v}_{\hat{\lambda}}^2 + \beta_{21}^2\hat{v}_{\hat{\omega}}^2)$ are approximately given by $d_F = (\hat{v}_{\hat{\lambda}}^2 + \beta_{21}^2\hat{v}_{\hat{\omega}}^2)^2/\{(\hat{v}_{\hat{\lambda}}^2)^2/(n_2 - 2) + (\beta_{21}^2\hat{v}_{\hat{\omega}}^2)^2/(n_1 - 2)\}$. We use $\hat{\beta}_{21}$ instead of β_{21} in the expression to obtain the estimated degrees of freedom $d_F^* = (\hat{v}_{\hat{\lambda}}^2 + \hat{\beta}_{21}^2\hat{v}_{\hat{\omega}}^2)^2/\{(\hat{v}_{\hat{\lambda}}^2)^2/(n_2 - 2) + (\hat{\beta}_{21}^2\hat{v}_{\hat{\omega}}^2)^2/(n_1 - 2)\}$. Then we have $a = \hat{\omega}^2 - t_{d_F^*,\alpha/2}^2\hat{v}_{\hat{\omega}}^2$, $b =$

Table 1

Confidence intervals for β_{21} in a simulation study with 2000 replications and true parameter values $(\beta_{10}, \beta_{20}, \beta_{21}, \omega) = (0.00, 0.00, 1.00, 1.00)$ for the linear regression model $Y_{1i} = \beta_{10} + X_{1i}\omega + \epsilon_{1i}; Y_{2j} = \beta_{20} + \beta_{21}X_{2j}\omega + \epsilon_{2j}$. "INL–LR" depicts the % of times that the interval by the likelihood ratio test was of infinite length, and "INL–FI" depicts the % of times that Fieller's interval was infinite length, either the entire real line or two infinite length disconnected intervals. Here the acronyms are IF, Inverse Fisher score method; HM, Hayya's method; NB, nonparametric bootstrap; PB, parametric bootstrap; FI, Fieller's interval; DIMER, Direct Integral Method for Ratios and LR Test—Likelihood ratio test.

	Mean of coverage			Mean of length			Median of length			90% Quantile of length		
Method	90% CI	95% CI	99% CI	90% CI	95% CI	99% CI	90% CI	95% CI	99% CI	90% CI	95% CI	99% CI
$n_1 = n_2 = 18, \mathrm{cv}(\hat{\omega}) = 0.26, \mathrm{cv}(\hat{\lambda}) = 0.26$; $\mathrm{mean}(\hat{\beta}_{10}, \hat{\beta}_{20}, \hat{\beta}_{21}, \hat{\omega}) = (0.01, 0.01, 1.10, 1.00), \mathrm{median}(\hat{\beta}_{10}, \hat{\beta}_{20}, \hat{\beta}_{21}, \hat{\omega}) = (0.01, 0.01, 1.00, 1.00)$												
IF	84.05	89.40	94.60	1.63	1.95	2.56	1.09	1.30	1.71	2.83	3.38	4.44
HM	88.50	92.90	96.70	1.74	2.08	2.73	1.15	1.37	1.80	2.31	2.75	3.61
NB	92.15	94.50	97.75	20.66	24.62	32.35	1.67	1.98	2.61	31.39	37.40	49.15
PB	92.00	94.20	97.35	38.84	46.28	60.83	1.49	1.78	2.34	22.75	27.10	35.62
FI	89.85	95.05	99.35	∞	∞	∞	1.39	1.80	3.08	4.28	8.25	∞
DIMER	91.45	95.90	99.50	2.69	4.92	63.53	1.43	1.88	3.35	3.74	6.12	37.32
LR Test	86.50	92.00	97.85	∞	∞	∞	1.27	1.65	2.78	3.72	6.50	∞
INL–LR	3.2%	6.6%	17.3%									
INL–FI	2.9%	5.7%	14.8%									
$n_1 = n_2 = 25, \mathrm{cv}(\hat{\omega}) = 0.21, \mathrm{cv}(\hat{\lambda}) = 0.21$; $\mathrm{mean}(\hat{\beta}_{10}, \hat{\beta}_{20}, \hat{\beta}_{21}, \hat{\omega}) = (-0.00, 0.00, 1.05, 1.00), \mathrm{median}(\hat{\beta}_{10}, \hat{\beta}_{20}, \hat{\beta}_{21}, \hat{\omega}) = (0.00, 0.00, 1.00, 1.00)$												
IF	86.15	92.15	96.50	1.35	1.60	2.11	0.95	1.13	1.49	2.17	2.59	3.41
HM	90.15	94.75	98.20	1.12	1.33	1.75	0.97	1.16	1.52	1.65	1.97	2.59
NB	92.55	95.30	98.45	10.25	12.21	16.05	1.17	1.40	1.83	7.55	8.99	11.82
PB	92.55	95.45	98.40	7.23	8.61	11.31	1.13	1.35	1.77	3.84	4.57	6.01
FI	90.15	95.90	99.60	∞	∞	∞	1.10	1.38	2.12	2.17	3.02	6.92
DIMER	91.15	96.30	99.70	1.78	2.75	10.96	1.12	1.42	2.23	2.20	3.06	6.74
LR Test	88.25	93.85	98.70	∞	∞	∞	1.04	1.31	1.99	2.02	2.79	6.15
INL–LR	0.6%	1.2%	5.9%									
INL–FI	0.5%	1.0%	4.5%									
$n_1 = n_2 = 50, \mathrm{cv}(\hat{\omega}) = 0.14, \mathrm{cv}(\hat{\lambda}) = 0.15$; $\mathrm{mean}(\hat{\beta}_{10}, \hat{\beta}_{20}, \hat{\beta}_{21}, \hat{\omega}) = (0.00, 0.00, 1.02, 1.00), \mathrm{median}(\hat{\beta}_{10}, \hat{\beta}_{20}, \hat{\beta}_{21}, \hat{\omega}) = (-0.00, 0.01, 1.00, 1.00)$												
IF	90.00	93.10	97.25	0.94	1.12	1.47	0.67	0.79	1.04	1.38	1.64	2.15
HM	90.65	95.50	98.65	0.71	0.84	1.11	0.67	0.79	1.04	0.95	1.13	1.48
NB	92.15	95.40	98.55	0.84	1.00	1.32	0.70	0.84	1.10	1.10	1.31	1.72
PB	92.15	96.10	98.65	0.80	0.95	1.25	0.71	0.84	1.11	1.09	1.29	1.70
FI	91.20	95.75	99.00	0.76	0.93	1.35	0.70	0.86	1.19	1.04	1.29	1.90
DIMER	91.50	95.80	99.10	0.77	0.94	1.36	0.71	0.87	1.21	1.05	1.31	1.94
LR Test	90.20	95.00	98.95	0.74	0.91	∞	0.69	0.83	1.16	1.01	1.26	1.85
INL–LR	0.0%	0.0%	0.1%									
INL–FI	0.0%	0.0%	0.0%									

$-2\hat{\omega}\hat{\lambda}$, and $c = \hat{\lambda}^2 - t^2_{d_F^*,\alpha/2}\hat{v}^2_{\hat{\lambda}}$ used in the **Supplementary Material**. Here, $\rho = 0$ since $\hat{\omega}$ and $\hat{\lambda}$ are independent.

3.4.3. Simulation results. Our simulations for model (2) compare the seven methods mentioned in Section 3.4.1. For simplicity, in all settings, we first fixed $v^2_{\epsilon_1} = v^2_{\epsilon_2} = 1$, and without loss of generality, let the intercepts β_{10} and β_{20} be 0. **Supplementary Material** Table S.4 gives some additional results when $v^2_{\epsilon_1} = v^2_{\epsilon_2} = 2$ and 3. We generated X_{1i} and X_{2j} independently from the standard normal distribution.

We considered two parameter configurations: $(\beta_{10}, \beta_{20}, \beta_{21}, \omega) = (0, 0, 1, 1)$, $(0, 0, 1, 0.75)$. For each parameter setting, we report simulation results for $(n_1, n_2) = (18, 18)$, $(25, 25)$, $(50, 50)$ with 2000 runs. In our experience in linear regression cases, and with these effect sizes, sample sizes higher than that typically lead to good numerical

Table 2

Confidence intervals for β_{21} in a simulation study with 2000 replications and true parameter values $(\beta_{10}, \beta_{20}, \beta_{21}, \omega) = (0.00, 0.00, 1.00, 0.75)$ for the linear regression model $Y_{1i} = \beta_{10} + X_{1i}\omega + \epsilon_{1i}$; $Y_{2j} = \beta_{20} + \beta_{21}X_{2j}\omega + \epsilon_{2j}$. "INL–LR" depicts the % of times that the interval by the likelihood ratio test was of infinite length, and "INL–FI" depicts the % of times that Fieller's interval was infinite length, either the entire real line or two infinite length disconnected intervals. Here the acronyms are IF, inverse Fisher score method; HM, Hayya's method; NB, nonparametric bootstrap; PB, parametric bootstrap; FI, Fieller's interval; DIMER, Direct Integral Method for Ratios; and LR Test, likelihood ratio test.

	Mean of coverage			Mean of length			Median of length			90% Quantile of length		
Method	90% CI	95% CI	99% CI	90% CI	95% CI	99% CI	90% CI	95% CI	99% CI	90% CI	95% CI	99% CI
$n_1 = n_2 = 18$, $\text{cv}(\hat{\omega}) = 0.35$, $\text{cv}(\hat{\lambda}) = 0.35$; $\text{mean}(\hat{\beta}_{10}, \hat{\beta}_{20}, \hat{\beta}_{21}, \hat{\omega}) = (0.01, 0.01, 2.85, 0.75)$, $\text{median}(\hat{\beta}_{10}, \hat{\beta}_{20}, \hat{\beta}_{21}, \hat{\omega}) = (0.01, 0.01, 1.00, 0.75)$												
IF	83.60	88.60	94.15	4.51	5.38	7.07	1.46	1.74	2.29	4.59	5.47	7.19
HM	86.45	91.65	95.55	2975	3544	4658	1.54	1.83	2.41	4.06	4.84	6.36
NB	93.35	95.10	97.75	74.05	88.24	116.0	4.54	5.41	7.11	105.1	125.2	164.6
PB	93.05	94.55	97.35	1634	1948	2559	3.55	4.23	5.56	94.25	112.3	147.6
FI	91.50	96.10	99.40	∞	∞	∞	2.13	2.97	7.75	∞	∞	∞
DIMER	92.80	96.55	99.55	7.57	15.87	56.16	2.15	3.05	8.28	10.03	25.88	105.1
LR Test	86.55	92.00	97.85	∞	∞	∞	1.92	2.70	7.84	∞	∞	∞
INL–LR	11.8%	18.8%	44.5%									
INL–FI	11.8%	18.2%	39.2%									
$n_1 = n_2 = 25$, $\text{cv}(\hat{\omega}) = 0.28$, $\text{cv}(\hat{\lambda}) = 0.28$; $\text{mean}(\hat{\beta}_{10}, \hat{\beta}_{20}, \hat{\beta}_{21}, \hat{\omega}) = (-0.00, 0.00, 1.17, 0.75)$, $\text{median}(\hat{\beta}_{10}, \hat{\beta}_{20}, \hat{\beta}_{21}, \hat{\omega}) = (0.00, 0.00, 1.00, 0.75)$												
IF	85.65	91.40	96.30	2.02	2.40	3.16	1.27	1.51	1.99	3.20	3.81	5.01
HM	89.45	94.15	97.75	5.06	6.03	7.92	1.30	1.55	2.03	2.69	3.20	4.21
NB	93.40	95.55	98.25	42.36	50.47	66.33	2.07	2.47	3.24	45.18	53.84	70.75
PB	93.10	95.50	98.30	53.39	63.62	83.61	1.91	2.28	2.99	35.70	42.54	55.90
FI	91.05	96.50	99.65	∞	∞	∞	1.59	2.11	3.90	5.72	15.03	∞
DIMER	92.40	96.95	99.75	4.53	9.82	35.54	1.62	2.16	4.15	4.64	7.96	61.76
LR Test	88.25	93.90	98.75	∞	∞	∞	1.49	1.96	3.65	5.14	11.50	∞
INL–LR	4.5%	8.0%	24.6%									
INL–FI	4.2%	7.1%	20.2%									
$n_1 = n_2 = 50$, $\text{cv}(\hat{\omega}) = 0.19$, $\text{cv}(\hat{\lambda}) = 0.20$; $\text{mean}(\hat{\beta}_{10}, \hat{\beta}_{20}, \hat{\beta}_{21}, \hat{\omega}) = (0.00, 0.00, 1.04, 0.75)$, $\text{median}(\hat{\beta}_{10}, \hat{\beta}_{20}, \hat{\beta}_{21}, \hat{\omega}) = (-0.00, 0.01, 1.00, 0.75)$												
IF	89.20	93.00	97.25	1.19	1.41	1.86	0.89	1.07	1.40	2.00	2.38	3.13
HM	90.80	95.30	98.45	0.99	1.17	1.54	0.89	1.06	1.39	1.43	1.70	2.23
NB	93.00	95.60	98.35	2.57	3.06	4.02	1.01	1.20	1.58	2.33	2.77	3.64
PB	92.75	96.10	98.65	3.79	4.52	5.93	1.00	1.19	1.57	2.19	2.61	3.43
FI	91.30	95.80	99.10	∞	∞	∞	0.97	1.21	1.77	1.73	2.28	4.13
DIMER	91.55	96.05	99.10	1.16	1.52	3.70	0.98	1.22	1.81	1.75	2.31	4.23
LR Test	90.20	95.00	98.95	∞	∞	∞	0.95	1.18	1.73	1.68	2.20	3.98
INL–LR	0.1%	0.4%	1.9%									
INL–FI	0.1%	0.2%	1.3%									

performances for all methods. Following Efron and Tibshirani (1994, p. 52), we used $B = 400$ bootstrap replications for all the bootstrap results reported in this article.

The results for the first parameter configuration $(\beta_{10}, \beta_{20}, \beta_{21}, \omega) = (0, 0, 1, 1)$ are given in Table 1 while Table 2 presents the results for setting $(\beta_{10}, \beta_{20}, \beta_{21}, \omega) = (0, 0, 1, 0.75)$. QQ plots (not shown here) comparing the quantiles of $\hat{\beta}_{21}$ with the quantiles of the standard normal distribution in the two parameter configurations with $n_1 = n_2 = 18$ clearly show that for small to moderate sample sizes, normal approximations are not appropriate.

Table 2 shows that when $n_1 = n_2 = 18$ the empirical mean of $\hat{\beta}_{21}$ is 2.85 when the true value is 1.00. The reason for this difference is that $\hat{\beta}_{21}$ follows a Cauchy like distribution, and one of characteristics for this distribution is that it has severely heavy tails. For example, the maximum of the absolute values of $\hat{\beta}_{21}$ reached $3,138$ over the 2000 runs in this case. Therefore, some severe outliers dramatically affected the empirical mean. In sharp contrast, the empirical median of $\hat{\beta}_{21}$ is 1.00.

Table 2 also displays the percentage of times the Fieller and likelihood ratio confidence intervals have infinite length.

The inverse Fisher information matrix algorithm has the lowest coverage probabilities. Hayya's method and the likelihood ratio test also have sub-nominal coverage probabilities when the sample sizes are small. Moreover, the latter has a positive probability to get infinite length. The performance of the two bootstrap methods is acceptable when the sample sizes are relatively large. When the sample sizes are small to moderate, the coverage rate of the bootstrap methods for the 90% confidence intervals are higher than 90%, while the coverage rate of the 99% confidence intervals is lower than 99%.

Fieller's interval has good performance overall in coverage. Here, we focus on cases that the sample sizes are small and moderate ($n_1 = n_2 = 18$ and $n_1 = n_2 = 25$), where Fieller's interval can be the real line or otherwise of infinite length. The inverse Fisher information method produced the shortest confidence interval lengths, not surprising, since its coverage rates are below the nominal values. Hayya's method remains stable but has a low coverage when the sample sizes are small. Compared with the two bootstrap methods, our method obviously has markedly shorter lengths in the 90% and 95% confidence intervals when the sample sizes are small and moderate, especially when $(n_1, n_2) = (18, 18)$. When the sample sizes are small, DIMER and Fieller's interval have similar median and interquartile ranges of lengths, but our method is much shorter in terms of mean and 90^{th} percentile of length.

4. Empirical Example and Further Simulations

4.1. *Method and Data Analysis*

The HEI-2005 and the NIH-AARP data available to us were described in Section 1. The sample sizes were 4300 males and 1916 females. Let $H(x) = \exp(x)/\{1 + \exp(x)\}$ be the logistic distribution function. Let $\ell = 1, 2$ denote men and women, respectively. Let $Y_{i\ell}$ denote the binary outcome of colorectal cancer for person $i = 1, \ldots, n_\ell$ in sample ℓ and let $X_{ij\ell}$ for $j = 1, \ldots, J = 12$ denote the HEI-2005 score for the jth dietary component. The traditional HEI-2005 analysis then posits a model $\mathrm{pr}(Y_{i\ell} = 1|X_{i1\ell}, \ldots, X_{iJ\ell}) = H(\alpha_\ell + \beta_\ell \sum_{j=1}^{J} X_{ij\ell})$, in other words, the HEI-scores are equally weighted. Notice here that the same predictor, $\sum_{j=1}^{J} X_{ij\ell}$, is to be used both for men and for women. In our case, we allow for the possibility that the predictor is the same in both populations, but the scores are weighted to be based on the data, so that our model is

$$\mathrm{pr}(Y_{i\ell} = 1|X_{i1\ell}, \ldots, X_{iJ\ell}) = H(\alpha_\ell + \beta_\ell \textstyle\sum_{j=1}^{J} \omega_j X_{ij\ell}), \quad (3)$$

where the weights $(\omega_1, \ldots, \omega_J)$ are estimated through the data. The model as such is not identified, but if we make the restriction that $\beta_1 = -1$, then it is identified: the negative value is because higher HEI-2005 scores, that is, better diets, lead to lower rates of colon cancer. Thus, with $\beta_1 = -1$ (3) becomes

$$\mathrm{pr}(Y_{i1} = 1|X_{i11}, \ldots, X_{iJ1}) = H(\alpha_1 - \textstyle\sum_{j=1}^{J} \omega_j X_{ij1});$$
$$\mathrm{pr}(Y_{i2} = 1|X_{i12}, \ldots, X_{iJ2}) = H(\alpha_2 + \beta_2 \textstyle\sum_{j=1}^{J} \omega_j X_{ij2}).$$

If we write $T_{i\ell} = \sum_{j=1}^{J} \omega_j X_{ij\ell}$, then we see that if the relative risk in men for changing $T_{i\ell}$ is R, the same change in women has a relative risk $R^{-\beta_2}$. Hence we wish to form a confidence interval for β_2. We fit model (3) by maximum likelihood, and the asymptotic covariance matrix Σ of $(\beta_2, \omega_1, \ldots, \omega_J)^{\mathrm{T}}$ was estimated using the Fisher information matrix.

To see how this relates to the Fieller problem, let $\omega = (\omega_1, \ldots, \omega_J)^{\mathrm{T}}$, $\lambda = \beta_2 \omega$, and e be the $J \times 1$ vector of ones. From Σ and the delta method, the asymptotic covariance matrix for $(\hat{\omega}, \hat{\lambda})$ can be constructed, and the covariance matrix of $(e^{\mathrm{T}}\lambda, e^{\mathrm{T}}\omega)$ is easily computed. Also, $\beta_2 = e^{\mathrm{T}}\lambda/e^{\mathrm{T}}\omega$ and $\hat{\beta}_2 = e^{\mathrm{T}}\hat{\lambda}/e^{\mathrm{T}}\hat{\omega}$. Thus, we see that $\hat{\beta}_2$ is the ratio of two asymptotically normal random variables, and hence DIMER, Fieller's method, etc. can be applied.

In the NIH-AARP study, the rate of colorectal cancer for men is 0.73%, while it is 0.48% for women. In the data analysis, we found that $\hat{\beta}_2 = -0.747$, so that if relative risk of 0.60 for men who improve their diet by a fixed amount, it is 0.68 for women who improve their diet the same amount. Thus, for colorectal cancer, the indication is that men are more susceptible, a well-known fact, and that they will have greater benefit for the same change in diet.

In the top panel of Table 3, we present the confidence intervals for the various methods. We see there that the confidence intervals for the inverse Fisher score method and Hayya's method are noticeably shorter than the others, the nonparametric bootstrap is quite a bit longer, and the parametric bootstrap, Fieller's interval, DIMER, and the likelihood ratio test are intermediate. The nonparametric bootstrap does not suggest differences in risk between men and women even at 90% confidence. With the exception of the nonparametric bootstrap, whose intervals we believe are much too long, see Section 4.2, all indications are that the risk for men and women for the same change in diet is statistically significant, with a p-value of < 0.01 for DIMER.

In the next subsection, we study whether the different lengths of the confidence intervals are reproducible in simulations, and through these simulations, which methods attained nominal coverage.

4.2. *Simulation*

The sample sizes were the same as in the data set, namely 4300 males and 1916 females. We used a bootstrap resample of the HEI-2005 scores in the NIH-AARP data as the covariates, separately for men and women, and generated 2000 data sets with binary outcome data according to the fit to the model (3), the parameter estimates of which are given in the caption to Table 3. The mean confidence intervals across the 2000 simulations are given in the bottom panel of Table 3. The result reflects the same phenomenon that was observed in the actual data set, namely the inverse Fisher score method and Hayya's method are noticeably shorter than the others, the nonparametric bootstrap is quite a bit

Table 3

Top Panel: confidence intervals for β_2 with the actual case-control data set of Section 4.1. The estimated values for parameter are $(\hat{\beta}_2, \hat{\alpha}_1, \hat{\alpha}_2) = (-0.747, -1.115, -1.024)$, $\hat{\omega} = (0.030, 0.018, 0.083, 0.033, -0.001, 0.081, 0.094, -0.043, 0.068, -0.020, 0.041, 0.098)^{\mathrm{T}}$. Bottom panel: average confidence intervals for β_2 in the simulation study of Section 4.2 for a logistic regression model with 2000 simulated case control data sets. Here the acronyms are IF, inverse Fisher score method; HM, Hayya's method; NB, nonparametric bootstrap; PB, parametric bootstrap; FI, Fieller's interval; DIMER, Direct Integral Method for Ratios; and LR Test, likelihood ratio test.

Data analysis			
Method	90% CI	95% CI	99% CI
IF	(−1.17, −0.33)	(−1.25, −0.25)	(−1.40, −0.09)
HM	(−1.18, −0.35)	(−1.26, −0.27)	(−1.41, −0.11)
NB	(−1.67, 0.18)	(−1.85, 0.35)	(−2.20, 0.70)
PB	(−1.26, −0.23)	(−1.36, −0.13)	(−1.56, 0.06)
FI	(−1.26, −0.33)	(−1.41, −0.24)	(−1.84, −0.02)
DIMER	(−1.26, −0.33)	(−1.41, −0.24)	(−1.84, −0.02)
LR Test	(−1.28, −0.30)	(−1.41, −0.22)	(−1.69, −0.05)
Simulation: Average confidence intervals			
Method	90% CI	95% CI	99% CI
IF	(−1.17, −0.37)	(−1.24, −0.29)	(−1.39, −0.15)
HM	(−1.18, −0.38)	(−1.26, −0.31)	(−1.40, −0.16)
NB	(−1.46, −0.08)	(−1.59, 0.06)	(−1.85, 0.32)
PB	(−1.26, −0.27)	(−1.36, −0.18)	(−1.55, 0.01)
FI	∞	∞	∞
DIMER	(−1.28, −0.35)	(−1.47, −0.22)	(−2.52, 0.53)
LR Test	(−1.28, −0.36)	(−1.40, −0.29)	(−1.67, −0.14)

longer, and the parametric bootstrap, DIMER, and the likelihood ratio test are intermediate. As seen in the previous simulations, the mean lengths of Fieller's interval in this case are infinite for 90%, 95%, and 99% intervals.

In Table 4, we show the confidence interval coverage performance of the various methods. The inverse Fisher score method and Hayya's method both have short confidence intervals generally, but also much less than nominal

Table 4

Analysis of the confidence intervals for β_2 in a simulation study of Section 4.2 for a logistic regression model with 2000 simulated data sets. There are 4300 males with 1075 individuals with colorectal cancer and 1916 females with 479 having the disease. "INL–LR" depicts the % of times that the interval by the likelihood ratio test was of infinite length, and "INL–FI" depicts the % of times that Fieller's interval was infinite length, either the entire real line or two infinite length disconnected intervals. Here the acronyms are IF, inverse Fisher score method; HM, Hayya's method; NB, nonparametric bootstrap; PB, parametric bootstrap; FI, Fieller's interval; DIMER, Direct Integral Method for Ratios; and LR Test, likelihood ratio test.

	Mean Coverage			Mean length			Median length			90% Quantile length		
Method	90% CI	95% CI	99% CI	90% CI	95% CI	99% CI	90% CI	95% CI	99% CI	90% CI	95% CI	99% CI
IF	84.10	91.65	97.20	0.79	0.95	1.24	0.76	0.91	1.19	0.99	1.18	1.55
HM	83.35	91.25	97.25	0.79	0.95	1.24	0.76	0.91	1.19	0.99	1.18	1.55
NB	96.55	98.45	99.60	1.38	1.65	2.17	1.19	1.41	1.86	2.08	2.48	3.26
PB	92.05	96.15	99.25	0.99	1.18	1.56	0.98	1.17	1.54	1.08	1.29	1.70
FI	87.60	94.30	99.00	∞	∞	∞	0.87	1.09	1.68	1.22	1.61	∞
DIMER	87.70	94.20	98.95	0.94	1.25	3.05	0.87	1.10	1.74	1.23	1.70	4.22
LR Test	86.85	92.80	98.25	0.92	1.11	1.53	0.87	1.05	1.43	1.18	1.44	2.01
INL–LR	0.0%	0.0%	0.0%									
INL–FI	1.6%	3.5%	10.4%									

coverage probability. The likelihood ratio test has longer intervals than the inverse Fisher score and Hayya's method, but it is still under coverage. The nonparametric bootstrap had by far the longest intervals, and here we see great over coverage. The parametric bootstrap, Fieller's method and DIMER have close to nominal coverage. For 95% confidence intervals, Fieller's method was of infinite length for almost 4% of the simulations. In this simulation, the parametric bootstrap performed somewhat better than DIMER, with its confidence intervals being somewhat shorter, although computationally it is, on average, 35 times slower to compute for data sets of this size.

For comparison purposes, the average computational time in these simulations for the Fisher Score, Hayya, nonparametric bootstrap, parametric bootstrap, Fieller's interval, DIMER, and likelihood ratio test were 0.07, 0.15, 75.94, 55.25, 0.15, 1.02, and 54.29 seconds, respectively. To do a more severe time test, we also generated cohort data similar in size to the NIH-AARP Study data (293, 615 males and 198, 245 females). For one such data set, the computational time for the six former methods (without the likelihood ratio test) was 9.80, 19.25, 8709.00, 3223.44, 19.25, and 20.34 seconds, respectively, indicating that the time of the parametric bootstrap was 159 times larger than that of DIMER.

5. Discussion

We have developed DIMER for constructing confidence intervals for the ratio of two location parameters. The method, based on analytical results and further approximations to account for nuisance parameters, is computationally fast. Our simulations indicated that compared with other methods in the literature, DIMER achieves coverage probabilities close to the nominal levels in all the different scenarios under consideration while providing competitive confidence interval lengths.

While we have no definitive explanation, it is a reasonable conjecture that an important reason why the DIMER method works well is that the distribution of the estimated ratio is heavy tailed. Our DIMER method appeared to be less affected by this problem due to its direct probability computation, although it is not unaffected, see below.

However, there are obvious cases that any of the intervals, including DIMER, may have poor performance. In particular, in the cases that Fieller intervals are of infinite length, we found in our simulations that DIMER intervals also increase in length, sometimes dramatically, especially when the p-value for testing the denominator being 0 or not is large. We found the same thing to happen to the other methods we have discussed: the results were poor, although in some cases better than DIMER. In the case of normality, only Fieller's interval is guaranteed to achieve its nominal coverage probability, at the potential cost of intervals of infinite length.

All the methods we have considered, other than Fieller's interval, are first-order correct, that is, their actual coverage probability is the nominal one $+O(n^{-1/2})$. There is a literature on second order correctness, that is, nominal level $+O(n^{-1})$, such as Laplace approximations, second order bootstrap, etc. It would be interesting to see how and whether these methods can be applied to our problem of finding a confidence interval for the ratio of two parameters. The properties of such methods such as confidence interval lengths and actual coverage in the settings we have considered are not at all clear.

6. Supplementary Material

The NIH-AARP Study of Diet and Health references in Section 4.1 can be accessed from the National Cancer Institute, but they require a proposal and a Material Transfer Agreement. The Supplementary Material includes a representative simulated data set for men and for women as in Section 4.2. The Supplementary Material referenced in Sections 2.1–2.5 and Sections 3.1–3.4 also includes a simulation when the slope and the intercept are dependent, a series of additional simulations including two with skew-normal regression errors, and definition of the HEI-2005 scores. Matlab programs implementing our method are also available at the *Biometrics* web site on Wiley Online Library.

Acknowledgements

This work was supported by a grant from the National Cancer Institute (R37-CA057030). We thank the patient associate editor and two patient referees for their useful comments.

References

Benton, D. and Krishnamoorthy, K. (2002). Performance of the parametric bootstrap method in small sample interval estimates. *Advances and Applications in Statistics* **2**, 269–285.

Beyene, J. and Moineddin, R. (2005). Methods for confidence interval estimation of a ratio parameter with application to location quotients. *BMC Medical Research Methodology* **5**, 1–7.

Davidian, M., Carroll, R. J., and Smith, W. (1988). Variance functions and the minimum detectable concentration in assays. *Biometrika* **75**, 549–556.

Efron, B. (1981). Nonparametric standard errors and confidence intervals. *The Canadian Journal of Statistics* **9**, 139–172.

Efron, B. and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*, 1st edition. New York, NY: Chapman and Hall/CRC Press.

Fieller, E. C. (1932). The distribution of the index in a bivariate normal distribution. *Biometrika* **24**, 428–440.

Fieller, E. C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society, Series B* **16**, 175–185.

Finney, D. J. (1978). *Statistical Methods in Biological Assay*, 3rd edition, London: Griffin.

Guenther, P. M., Reedy, J., and Krebs-Smith, S. M. (2008). Development of the Healthy Eating Index-2005. *Journal of the American Dietetic Association* **108**, 1896–1901.

Hayya, J., Armstrong, D., and Gressis, N. (1975). A note on the ratio of two normally distributed variables. *Management Science* **21**, 1338–1341.

Hubert, J. J. (1984). *Bioassay*, 2nd edition, 26–39. Dubuque: Kendall-Hunt.

Pham-Gia, T., Turkkan, N., and Marchand, E. (2006). Density of the ratio of two normal random variables and applications. *Communications in Statistics: Theory and Methods* **35**, 1569–1591.

Redmond, C. K. (2005a). Parallel line assay. *Encyclopedia of Biostatistics*. New York: Wiley. DOI: 10.1002/0470011815.b2a06020.

Redmond, C. K. (2005b). Radioimmunoassay. *Encyclopedia of Biostatistics*. New York: Wiley. DOI: 10.1002/0470011815.b2a06021.

Redmond, C. K. (2005c). Slope-ratio assay. *Encyclopedia of Biostatistics*. New York: Wiley. DOI: 10.1002/0470011815.b2a06023.

Reedy, J., Mitrou, P. N., Krebs-Smith, S. M., Wirfält, E., Flood, A., Kipnis, V., Leitzmann, M., Mouw, T., Hollenbeck, A., Schatzkin, A., and Subar, A. F. (2008). Index-based dietary patterns and risk of colorectal cancer: The NIH-AARP Diet and Health Study. *American Journal of Epidemiology* **168**, 38–48.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin* **2**, 110–114.

Sherman, M., Maity, A., and Wang, S. (2011). Inferences for the ratio: Fieller's interval, log ratio, and large sample based confidence intervals. *AStA Advances in Statistical Analysis*, **95**, 313–323.

Welch, B. L. (1947). The generalization of "student's" problem when several different population variances are involved. *Biometrika* **34**, 28–35.

Received November 2013. *Revised December* 2014.
Accepted February 2015.