Krishna K. Saha*, Daniel Miller and Suojin Wang

A Comparison of Some Approximate Confidence Intervals for a Single Proportion for Clustered Binary Outcome Data

DOI 10.1515/ijb-2015-0024

Abstract: Interval estimation of the proportion parameter in the analysis of binary outcome data arising in cluster studies is often an important problem in many biomedical applications. In this paper, we propose two approaches based on the profile likelihood and Wilson score. We compare them with two existing methods recommended for complex survey data and some other methods that are simple extensions of well-known methods such as the likelihood, the generalized estimating equation of Zeger and Liang and the ratio estimator approach of Rao and Scott. An extensive simulation study is conducted for a variety of parameter combinations for the purposes of evaluating and comparing the performance of these methods in terms of coverage and expected lengths. Applications to biomedical data are used to illustrate the proposed methods.

Keywords: beta-binomial, clustered binary data, confidence interval, coverage, expected length

1 Introduction

Binary outcome data sampled from clusters arise frequently in many biomedical, toxicological, clinical medicine, and epidemiological applications. The observed binary outcome data often exhibit greater or lesser variability than that predicted by a simple binomial model, referred to as over/under-dispersion [1]. There are several reasons that may lead to the over/under-dispersion in binary data. For instance, in a case-controlled study of familial aggregation of chronic obstructive pulmonary disease (COPD) [2], siblings within each family are correlated and the number of impaired pulmonary function (IPF) cases per family may be over-dispersed as compared to a binomial model. The 95% profile based confidence interval for intraclass correlation proposed by Saha ([3]) using the data in the study mentioned above (see Table 5 of Liang *et al.* [2]) is (0.0593, 0.4006), which supports the significance of within-family correlation. As a result, it shows that the observed variance 0.1418 in the estimated proportion of the IPF cases per family is 1.38 times larger than the predicted variance 0.1026 obtained using a binomial model. This concludes that standard approaches (see, for example, Hogg and Tanis [4], pp. 308–310) of analyzing such data that ignore the cluster structure may result in underestimation of the true standard error of the estimated infected rate when the correlation between siblings per family is positive. Furthermore, inference methods concerning the parameters of interest based on the binomial model in such data may significantly inflate the Type I error rate [5]. Although a number of confidence intervals for a single proportion based on clustered data have been studied for complex survey data, little attention has been paid to model-based approaches for inferring about the proportion in the analysis of clustered binary data. Kleinman [6] studied the properties of the maximum likelihood (ML) and the method of moments (MM) estimators for the proportion parameter based on parametric and semiparametric models for clustered binary data. Based on several model structures, Paul and Islam [7] investigated the joint estimation of the proportion and dispersion parameters in terms of bias and efficiency. Surprisingly, these approaches were not extended to investigate the coverage probabilities of confidence interval estimation of the proportion.

*Corresponding author: Krishna K. Saha, Department of Mathematical Sciences, Central Connecticut State University, 1615 Stanley Street, New Britain, CT 06050, USA, E-mail: sahakrk@ccsu.edu

Daniel Miller, Department of Mathematical Sciences, Central Connecticut State University, 1615 Stanley Street, New Britain, CT 06050, USA

Suojin Wang, Department of Statistics, Texas A&M University, College Station, TX 77843, USA

Developing a confidence interval for the proportion parameter using clustered binary data is an important problem. For example, in the diagnostic accuracy of contrast-enhanced multi-detector row spiral computed tomography coronary angiography [8], interval estimates of sensitivity (true positive rate) and specificity (true negative rate) are often used at the patient level, at the coronary artery level, and at the coronary artery segment level. Another example involves the estimation of sensitivity and specificity to assess the accuracy of radiologists' readings in a mammogram screening study [9], where the proportion of positive readings in cancer cases and the proportion of negative readings in non-cancer cases within each radiologist may be over-dispersed. To make inferences on sensitivity and specificity, one usually uses the inference methodology developed for a single proportion. There is an abundance of literature pertaining to inferences for a single proportion based on non-clustered binary data (for example, Agresti and Coull [10] and Newcombe [11]). However, little attention has been paid to such inferences in the case of small to moderate sample size clustered binary data using parametric and semiparametric models. For complex survey data, some authors have developed alternative methods extending those derived for non-clustered binary data. In particular, the modified Clopper-Pearson (MCP) method and the modified Wilson score (MWS) method are recommended for analyzing complex survey data (see Korn and Graubard [12], page 65). However, in some situations the MCP method is somewhat conservative while the MWS method shows lack of coverage when the variability of the weights is small or large. Moreover, the sampling weights are not readily available in biomedical applications though they are required to calculate the effective sample size in order to find these intervals.

Lui [13] derived three methods based on model assumptions using the estimation of the intraclass correlation by analysis of variance, but the performances of these methods were not examined. Rutter [14] introduced bootstrap interval methods for sensitivity and specificity to measure the diagnostic accuracy with patient-clustered data using bootstrapping to estimate the variance, but these methods are computationally demanding. For the balanced data set-up, Kim and Lee [9] proposed an asymptotic confidence interval for a single proportion based on the beta-binomial distribution which works well for larger proportions when the cluster sizes are over 25, but suffers from serious under-coverage for small numbers of clusters. In many practical problems, the cluster sizes are often not equal and small (< 15) or the number of clusters is small to moderate (see, for example, Zhou *et al.* [8], page 112). In order to assess the accuracy of computer-aided detection enhanced computed tomography colonography for the detection of polyps, Zhou *et al.* [8] obtained the asymptotic confidence interval for sensitivity of clustered binary data using a ratio estimator for the variance given by Rao and Scott [5]. However, this method shows serious under-coverage (see, Figure 1 and Paul and Zaihra [15], p. 4219).

The main focus of this paper is to develop asymptotic confidence intervals for a single proportion arising in cluster studies. In particular, in Section 3 we propose two new approaches based on the Wilson score and profile likelihood that will properly incorporate the intraclass correlation structure. In addition, we consider a number of extensions of existing methods in order for them to be feasible for clustered binary outcome data. Section 4 conducts a simulation study to assess the performances of these intervals in comparison with two existing methods recommended for analyzing survey data in terms of coverage and interval length. The methods developed in this paper are applied to analyze medical data sets in Section 5. Some concluding remarks are given in Section 6.

2 Models

2.1 Beta-binomial model

Let x_i ($i = 1, \dots, k$) be the number of individuals affected by the risk factor among the n_i individuals. Suppose that x_i is a conditional binomial response variable with parameter p_i and denominator n_i , where p_i is itself a random variable following a beta distribution with mean π and variance $\pi(1 - \pi)\phi$. The unconditional distribution of x_i is the beta-binomial distribution with the probability mass function, denoted by $BB(\pi, \phi)$, given by

Figure 1: The observed coverage probability and the expected interval length of 95% nominal confidence intervals for the proportion π based on the 15 methods discussed in Section 3. Each box plot was constructed based on 180 parameter combinations.

$$f(x_i|\pi, \phi) = \binom{n_i}{x_i} \frac{\prod_{j=0}^{x_i-1} [(1-\phi)\pi + j\phi] \prod_{j=0}^{n_i-x_i-1} [(1-\pi)(1-\phi) + j\phi]}{\prod_{j=0}^{n_i-1} [(1-\phi) + j\phi]} \quad (1)$$

for $x_i = 0, 1, 2, \dots, n_i$. As $\phi \rightarrow 0$, $BB(\pi, \phi)$ becomes a simple binomial model with parameters n_i and π . Note that the beta-binomial distribution has been widely used for biomedical data (see, for example, Williams [16]). Due to its simplicity, most authors have considered this model. Moreover, this model is the most sensitive to the departure from the binomial model and is a superior model compared to its competitive models for the analysis of correlated binomial data [1].

2.2 Semi-parametric models

In some situations, the full parametric assumption may be too restrictive in which case a more flexible model can be used that only specifies the mean and variance of the data distribution (see Paul and

Islam [17]). Let n_i ($i=1, \dots, k$) be the number of individuals in the i th cluster who are exposed to a risk factor. Let x_{ij} ($j=1, \dots, n_i$) be the binary outcome of the j th individual of the i th cluster so that the probability of the j th individual in the i th cluster being affected by the risk factor, $p_i = P(x_{ij} = 1)$ and the interclass correlation between individuals in the i th cluster, $\phi = \text{corr}(x_{ij}, x_{ij'}), j \neq j'$. Then $x_i = \sum_{j=1}^{n_i} x_{ij}$ is the total number of the n_i individuals affected by the risk factor. Suppose that $x_i (i=1, \dots, k)$ is a sampled value of random variable X_i from a population such that the expected value of the number of affected individuals $E(X_i) = n_i\pi$ and the variance of the number of affected individuals $\text{Var}(X_i) = n_i\pi(1-\pi)\{1 + (n_i-1)\phi\}$, where π is the expected proportion of affected individuals in the population and ϕ measures the correlation within the same cluster, referred to as the intraclass correlation coefficient. This specification of the mean and variance coincides with the extended BB model. This model includes several special cases that allow the data to come from any distribution that is specified by only the first two moments of the binomial response with some unknown common intraclass correlation. More specifically, the data come from any distribution that belongs to the family of distributions having mean $n_i\pi$, variance proportional to $n_i\pi(1-\pi)$, and support on the integers $(0, 1, \dots, n_i)$. Note that the variance structure of the form $\text{Var}(X_i) = n_i\pi(1-\pi)\{1 + (n_i-1)\phi\}$ is the most popular and is generally robust against variance misspecification [18].

3 Confidence intervals for the single proportion

Since we will deal with and compare a large number of methods for constructing confidence intervals (CI), we now list all of them and their acronyms for easy reference:

PL: profile likelihood	EQL: extended quasi-likelihood
WI ₁ and WI ₂ : two versions of the Wilson score modified for clustered binary data	DEQL: double extended quasi-likelihood
WA ₁ and WA ₂ : two versions of the Wald CI	QEE: quadratic estimating equations
R ₁ and R ₂ : two versions of the ratio estimator	MCP: modified Clopper-Pearson method for complex survey data
G ₁ and G ₂ : two versions of the generalized estimating equation	MWS: modified Wilson score for complex survey data
ML: maximum likelihood	

3.1 The CI based on PL

Altaye *et al.* [19] discussed that an obvious approach to constructing a confidence interval for the parameter of interest may not perform well with extreme true values or when the sample size is small. Here we use a profile likelihood based confidence interval approach which has been shown to provide accurate results when computing confidence limits for a single proportion [11] or the difference between two proportions [20] in the case of non-clustered binary data. Let $l(\pi, \phi)$ be the log-likelihood function, where π is the parameter of interest and ϕ is the nuisance parameter. Also, let $l_p(\pi) = l(\pi, \hat{\phi}(\pi))$ be the profile likelihood for π , where $\hat{\phi}(\pi)$ is obtained from the reduced model with respect to ϕ keeping π fixed. Then the approximate $100(1-\alpha)\%$ profile likelihood (PL) based confidence interval for π is given by

$$\left\{ \pi : l_p(\pi) \geq l(\hat{\pi}, \hat{\phi}) - \frac{1}{2}\chi_{1,\alpha}^2 \right\},$$

where $\hat{\pi}$ and $\hat{\phi}$ are the estimates of π and ϕ in the full model and $\chi_{1,\alpha}^2$ is the $100(1-\alpha)$ percentile of a chi-squared distribution with one degree of freedom. As discussed in Section 2, the beta-binomial is a superior model for clustered binary data compared to its competitive models. We consider the beta-binomial model for clustered binary data in order to obtain the PL based confidence interval for π . Moreover, the estimates

of π and ϕ for the beta-binomial model can be obtained following the procedure discussed in the Supplementary Materials. Finally, the endpoints of the confidence interval can be obtained by solving the system of nonlinear equations following the methodology introduced by Venzon and Moolgavkar (see, for example, Pradhan *et al.* [20]). Alternatively, the interval limits can be obtained by finding the two roots of the above equation, one in the interval $(0, \hat{\pi})$ and the other in the interval $(\hat{\pi}, 1)$, using either the bisection method or Brent's method.

3.2 The Wilson score interval

From the above semi-parametric model, an estimator of π can easily be obtained as the overall sample proportion $\hat{\pi} = x/n$, where $x = \sum_i^k x_i$ and $n = \sum_i^k n_i$. The variance of $\hat{\pi}$ is given by $\text{Var}(\hat{\pi}) = \pi(1-\pi)\xi/n$, where $\xi = \sum n_i[1 + (n_i - 1)\phi]/n$. Using the central limit theorem, it can be shown that $n^{1/2}(\hat{\pi} - \pi)/\sqrt{\pi(1-\pi)\xi}$ converges in distribution to the standard normal distribution as $k \rightarrow \infty$, where $\hat{\xi}$ is obtained by replacing ϕ by its estimate $\hat{\phi}$. Then the approximate $100(1-\alpha)\%$ Wilson confidence interval for π is the roots of the quadratic equation

$$P(n(\hat{\pi} - \pi)^2 / [\pi(1-\pi)\hat{\xi}] \leq z_{\alpha/2}^2) = 1 - \alpha.$$

After some straightforward algebra, it can be obtained as

$$\tilde{\pi} \pm \frac{z_{\alpha/2}}{\tilde{n}} \sqrt{n\hat{\pi}(1-\hat{\pi})\hat{\xi} + \frac{\hat{\xi}^2 z_{\alpha/2}^2}{4}},$$

where

$$\tilde{\pi} = \frac{n\hat{\pi} + 0.5\hat{\xi}z_{\alpha/2}^2}{n + \hat{\xi}z_{\alpha/2}^2} = \frac{x + 0.5\hat{\xi}z_{\alpha/2}^2}{n + \hat{\xi}z_{\alpha/2}^2}, \quad \tilde{n} = n + \hat{\xi}z_{\alpha/2}^2,$$

and $z_{\alpha/2}$ is the $100(1-\alpha/2)$ th percentile of the standard normal distribution. It is worthwhile to note here that for non-clustered data when there is no cluster effect, that is, $\phi = 0$ (or $\xi = 1$) the same intervals are produced (see, for example, Newcombe [11]). The estimate $\hat{\phi}$ can be obtained using the ML method based on the beta-binomial model discussed in Section 2.1 as well as the analysis of variance (ANOVA) method used by Paul and Zaihra [15]. The ANOVA-type estimate of ϕ is given by $\hat{\phi}_a = (\text{BMS} - \text{WMS})/[\text{BMS} + (n^* - 1)\text{WMS}]$, where $\text{BMS} = [\sum_i x_i^2/n_i - (\sum_i x_i)^2/\sum_i n_i]/(k-1)$ and $\text{WMS} = [\sum_i x_i - \sum_i x_i^2/n_i]/\sum_i (n_i - 1)$ are the between mean-squared and within mean-squared errors, respectively, and $n^* = [(\sum_i n_i)^2 - \sum_i n_i^2]/[(k-1)\sum_i n_i]$. Therefore, one can obtain Wilson CIs for π using the above interval (π_l, π_u) by substituting ANOVA and ML estimates of ϕ in the equation for $\hat{\xi}$ above. We denote the respective intervals as WI_1 and WI_2 .

3.3 Extensions of other methods

We now consider some additional approaches that extend existing methods:

The Wald CIs: From the above, we see that the sample proportion $\hat{\pi} = x/n$ is an unbiased estimator of π with the variance of $\hat{\pi}$ given by $\text{Var}(\hat{\pi}) = \pi(1-\pi)\xi/n$. Then, as $k \rightarrow \infty$, $\hat{\pi}$ follows the normal distribution with mean π and variance $\pi(1-\pi)\xi/n$. The resulting approximate $100(1-\alpha)\%$ Wald CI for π is given by $\hat{\pi} \pm z_{\alpha/2} \sqrt{\hat{\pi}(1-\hat{\pi})\hat{\xi}/n}$. Similar to the Wilson CIs, we also obtain two versions of the Wald CI, WA_1 and WA_2 , for π using the ANOVA and ML estimates of ϕ in the equation for $\hat{\xi}$ above, respectively.

RE based CIs: Following the result provided by Rao and Scott [5], one can obtain the corrected estimated variance of $\hat{\pi}$ by

$$\vartheta_R = \frac{k}{(k-1)n^2} \sum_{i=1}^k (x_i - n_i \hat{\pi})^2.$$

Note that ϑ_R is a consistent estimator of the $\text{Var}(\hat{\pi})$. In addition, as $k \rightarrow \infty$, $\hat{\pi}$ is asymptotically $N(\pi, \vartheta_R)$. Then the approximate $100(1-\alpha)\%$ confidence interval for π based on RE is given by $\hat{\pi} \pm z_{\alpha/2} \sqrt{\vartheta_R}$. Based on this approach, Paul and Zaihra [15] developed an interval estimate of the risk difference and replaced $\hat{\pi}$ by $\hat{\pi}^* = (x_i + 0.5)/(n_i + 1)$ in ϑ_R irrespective of whether $\hat{\pi}$ is 0 or 1. We also replace $\hat{\pi}$ by $\hat{\pi}^*$ in the equation for ϑ_R above and denote the resulting confidence interval as R_1 . For non-clustered data, Agresti and Coull [10] have adjusted the RE intervals for a single binomial proportion by using $\hat{\pi}^c = (x_i + c)/(n_i + 2c)$ with $c = 2$ for π . Replacing $\hat{\pi}$ with this $\hat{\pi}^c$ results in an interval denoted by R_2 .

GEE based CIs: Paul and Zaihra [15] applied the generalized estimating equation (GEE) approach of Zeger and Liang [21] to basic binary data and obtained an estimate of a proportion from clustered correlated binary data and a sandwich estimate of its variance, which are given by

$$\hat{\pi} = x_i/n_i \text{ and } V_G = \frac{\sum_{i=1}^k (x_i - n_i \hat{\pi})^2}{n_i^2},$$

respectively. As $k \rightarrow \infty$, it follows that $\hat{\pi}$ is asymptotically $N(\pi, V_G)$. Then the approximate $100(1-\alpha)\%$ confidence interval for π is given by $\hat{\pi} \pm z_{\alpha/2} \sqrt{V_G}$. Similar to R_1 and R_2 , we replace $\hat{\pi}$ by $\hat{\pi}^*$ and $\hat{\pi}^c$ resulting in two CIs denoted by G_1 and G_2 , respectively.

ML based CI: The log-likelihood of the beta-binomial model, apart from a constant, can be written as

$$l(\pi, \phi) = \sum_{i=1}^k \left[\sum_{r=0}^{x_i-1} \ln\{(1-\phi)\pi + r\phi\} + \sum_{r=0}^{n_i-x_i-1} \ln\{(1-\pi)(1-\phi) + r\phi\} - \sum_{r=0}^{n_i-1} \ln\{(1-\phi) + r\phi\} \right].$$

The ML estimators $\hat{\pi}_{ml}$ and $\hat{\phi}_{ml}$ of π and ϕ and the asymptotic variance $\text{Var}(\hat{\pi}_{ml})$ can be obtained using $l(\pi, \phi)$. Further details are provided in the Supplementary Materials. Using the asymptotic property of the ML estimate, we obtain an approximate $100(1-\alpha)\%$ confidence interval for π as $\hat{\pi}_{ml} \pm z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\pi}_{ml})}$, where $\widehat{\text{Var}}(\hat{\pi}_{ml})$ is the estimated variance of $\hat{\pi}_{ml}$ obtained from the Supplementary Materials by replacing the parameters π and ϕ by $\hat{\pi}_{ml}$ and $\hat{\phi}_{ml}$, respectively.

EQL based CI: Paul and Saha [22] obtained the EQL using the mean and variance of x_i specified in the semi-parametric model above which, apart from a constant, is

$$Q^+ = \sum_{i=1}^k \left[-\frac{1}{2} \ln\{1 + (n_i - 1)\phi\} + \frac{x_i \ln\left(\frac{n_i \pi}{x_i}\right) + (n_i - x_i) \ln\left(\frac{n_i(1-\pi)}{n_i - x_i}\right)}{1 + (n_i - 1)\phi} \right].$$

Now, using Q^+ one can obtain the EQL estimates $\hat{\pi}_{eql}$ and $\hat{\phi}_{eql}$ and the asymptotic variance $\text{Var}(\hat{\pi}_{eql})$ (see further details in the Supplementary Materials). Following the results of Inagaki [23], $\hat{\pi}_{eql}$ is asymptotically $N(\pi, \text{Var}(\hat{\pi}_{eql}))$ as $k \rightarrow \infty$. Then an approximate $100(1-\alpha)\%$ confidence interval of π based on EQL is given by $\hat{\pi}_{eql} \pm z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\pi}_{eql})}$, where $\widehat{\text{Var}}(\hat{\pi}_{eql})$ is the estimated variance of $\hat{\pi}_{eql}$ obtained from the Supplementary Materials after replacing the parameters π and ϕ by $\hat{\pi}_{eql}$ and $\hat{\phi}_{eql}$, respectively.

DEQL based CI: Based on the semi-parametric model, we obtain the profile double extended quasi-likelihood from Paul and Saha [22], which, apart from a constant, is

$$\begin{aligned}
p_V(Q) = & \sum_{i=1}^k \left[\left(x_i + \frac{\pi}{\delta} - \frac{1}{2} \right) \ln \left(x_i + \frac{\pi}{\delta} \right) + \left(n_i - x_i + \frac{1-\pi}{\delta} - \frac{1}{2} \right) \ln \left(n_i - x_i + \frac{1-\pi}{\delta} \right) \right. \\
& - \left(n_i + \frac{1}{\delta} - \frac{1}{2} \right) \ln \left(n_i + \frac{1}{\delta} \right) + \frac{\delta}{12(\pi + \delta x_i)} + \frac{\delta}{12\{1 - \pi + \delta(n_i - x_i)\}} - \frac{\delta}{12(1 + \delta n_i)} \\
& - \left(\frac{\pi}{\delta} + \frac{1}{2} \right) \ln \left(\frac{\pi}{\delta} \right) - \left(\frac{1-\pi}{\delta} - \frac{1}{2} \right) \ln[(1-\pi)/\delta] + \left(\frac{1}{\delta} + \frac{1}{2} \right) \ln(1/\delta) - \frac{\delta}{12\pi} \\
& \left. - \frac{\delta}{12(1-\pi)} + \frac{\delta}{12} \right],
\end{aligned}$$

with $\delta = \phi(1 - \phi)^{-1}$. The DEQL estimates $\hat{\pi}_{de}$ and $\hat{\delta}_{de}$ and the asymptotic variance $\text{Var}(\hat{\pi}_{de})$ can be obtained following the procedure discussed in the Supplementary Materials. Similar to the EQL procedure, $\hat{\pi}_{de}$ is also asymptotically $N(\pi, \text{Var}(\hat{\pi}_{de}))$ as $k \rightarrow \infty$. Then a $100(1-\alpha)\%$ confidence interval of π is $\hat{\pi}_{de} \pm z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\pi}_{de})}$, where $\widehat{\text{Var}}(\hat{\pi}_{de})$ is the estimated variance of $\hat{\pi}_{de}$ obtained from the Supplementary Materials after replacing the parameters π and ϕ by $\hat{\pi}_{de}$ and $\hat{\delta}_{de}$, respectively.

QEE based CI: The QEE estimates $\hat{\pi}_{qee}$ and $\hat{\phi}_{qee}$ are the solutions to the optimal quadratic estimating equations (QEE) (see, for example, Paul and Islam [7]) for the parameters π and ϕ obtained based on the above semiparametric model. Also, the sandwich variance of $\hat{\pi}_{qee}$ can be obtained using the results of Inagaki [23]. The detailed derivation about the estimates and their sandwich variances are provided in the Supplementary Materials. As $k \rightarrow \infty$, it can be shown that $\hat{\pi}_{qee}$ is asymptotically $N(\pi, \text{Var}(\hat{\pi}_{qee}))$. Then an approximate $100(1-\alpha)\%$ confidence interval for π is given by $\hat{\pi}_{qee} \pm z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\pi}_{qee})}$, where $\widehat{\text{Var}}(\hat{\pi}_{qee})$ is the estimated variance of $\hat{\pi}_{qee}$ obtained from the Supplementary Materials after replacing the parameters π and ϕ by $\hat{\pi}_{qee}$ and $\hat{\phi}_{qee}$, respectively.

4 Simulation studies

[illegible]

details), quoted by Kupper *et al.* [25], for $k = 20, 50, 100$. Ten values of $\pi = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5$, and three values of $\phi = 0.1, 0.3, 0.5$, were considered in our investigation. However, we only present representative results for brevity. Following Paul and Islam [17], 10,000 data sets were generated from the beta-binomial distribution. Following Paul and Saha [22], we also considered negative estimates of ϕ in our simulation, with the restriction that $\hat{\phi} > -1/(n_{\max} - 1)$, where $n_{\max} = \max\{n_i, i = 1, \dots, k\}$. The observed coverage probability (CP) and the expected interval length (EL) for two-sided confidence intervals (l_t, u_t) for π were obtained by

$$CP = \frac{\sum_{t=1}^{10,000} I(l_t \leq \pi \leq u_t)}{10,000} \text{ and } EL = \frac{\sum_{t=1}^{10,000} (u_t - l_t)}{10,000},$$

where $I = 1$ if $l_t \leq \pi \leq u_t$, and $I = 0$, otherwise.

The coverage properties and expected lengths for all the methods are almost identical for both the 90% and 95% pre-assigned confidence levels. We present simulation results for only a pre-assigned confidence level of 95% with the results for a 90% nominal confidence level being provided in the Supplementary Materials (see, Figures W2–W6 and Tables W5–W8). The box plots and the medians of the observed CPs and the ELs obtained from the 15 interval procedures for the 180 (6 values of k , 10 values of π , and 3 values of ϕ) parameter combinations are reported in Figure 1 and Table 1, respectively. The horizontal lines in the coverage probability figure indicate the CPs 0.94, 0.95, and 0.96, respectively. From Figure 1 and Table 1, it can be seen that the median CPs for all methods are, in general, close to the pre-assigned confidence level, except for EQL, DEQL, G_1 , and MCP. The EQL, DEQL, and MCP intervals are conservative as their median CPs lie above 96%, while the G_1 intervals are liberal as evidenced by median CPs below 94%. These results are clearly seen in Figure 1 where almost 75% of the CPs for EQL, DEQL, and MCP lie above the horizontal line at 0.96, and a similar proportion of the CPs for G_1 lie below the horizontal line at 0.94. Although G_1 and R_1 have liberal behaviors in coverage (also reported by Paul and Zaihra [15], p. 4219), their adjusted versions, G_2 and R_2 , clearly show much improvement in coverage as most of their CPs lie between the horizontal lines at 0.94 and 0.96. The remaining eight procedures: ML, QEE, WA_1 , WA_2 , WI_1 , WI_2 , PL, and MWS, perform reasonably well. However, in some situations, ML, QEE, WA_1 , and WA_2 tend to be liberal (i.e., CPs < 0.94). Based on coverage, overall the WI_1 , WI_2 , PL, and MWS intervals outperform the other intervals. As expected, the ELs for EQL, DEQL, and MCP are wider than those for the other intervals considered here. In general, the ELs for ML and PL are shortest, while the ELs for the remaining nine intervals are very similar except for QEE. However, WI_2 and WA_2 have a slight edge in EL compared to WI_1 and WA_1 .

Table 1: Median coverage probability (CP) and median expected length (EL) of the 95% confidence intervals for π based on 180 parameter combinations for 15 methods.

Method	Median CP	Median EL	Length comparison individual/ML
PL	0.943	0.156	0.971
WI_1	0.950	0.167	1.043
WI_2	0.951	0.164	1.025
WA_1	0.943	0.168	1.051
WA_2	0.942	0.165	1.031
R_1	0.941	0.166	1.039
R_2	0.944	0.166	1.039
G_1	0.937	0.165	1.030
G_2	0.942	0.165	1.030
ML	0.941	0.160	1.000
EQL	0.974	0.197	1.227
DEQL	0.968	0.191	1.195
QEE	0.944	0.188	1.175
MCP	0.965	0.179	1.119
MWS	0.955	0.170	1.063

We now evaluate the methods stratified by different parameter space points. It has been seen from the above results that EQL, DEQL, and MCP are somewhat conservative, and the CPs for G_2 and R_2 are better than those for G_1 and R_1 . However, it shows that MCP has a less conservative property than those of EQL and DEQL. In the interest of brevity, we present the observed CP and EL results only for the 11 generally more competitive methods, namely, PL, WI_1 , WI_2 , WA_1 , WA_2 , R_2 , G_2 , ML, QEE, MCP, and MWS. Since the results for fixed cluster sizes between $k=19$ and $k=27$ as well as for ED cluster sizes between $k=20$ and $k=30$ are almost identical in most data situations, in Figures 2–5 we present the observed CPs and the ELs only for fixed cluster sizes of $k = 19$ and 73 and for ED cluster sizes of $k = 20$ and 50 . In addition to Figures 2–5, the simulation results for selected parameter combinations are also presented in Tables 2 and 3 to clearly show how the results of CPs and ELs are affected by different subsets of π and cluster sizes.

Figure 2: The observed coverage probability of 95% nominal confidence intervals for the proportion π based on the 11 methods for fixed litter sizes. Each box plot was constructed based on 10 parameter combinations.

Figure 3: The observed coverage probability of 95% nominal confidence intervals for the proportion π based on the 11 methods for ED litter sizes. Each box plot was constructed based on 10 parameter combinations.

The results in Figures 2 and 3 and Table 2 show that, irrespective of fixed or variable cluster sizes, the coverage properties of all the methods are very similar. As expected, the variations of CPs for different values of π for all the methods decrease as k increases (see, for example, Figure 2(iii) for $k = 19$ and Figure 2(iv) for $k = 73$). Although the median coverage probability for ML is between 94% and 95%, we see inconsistent coverage in some situations. For instance, when $\pi = 0.05$, the CPs are much larger than the nominal level (see, for example, Table 2 when $k = 19$ and $k = 20$), and for moderate π 's, the CPs are much smaller than the nominal level (see, for example, Table 2 when $k = 19$, $\phi = 0.5$, and $0.15 \leq \pi \leq 0.35$). Like the ML intervals, the QEE method shows inconsistency in coverage when $\pi < 0.15$. However, in most data situations the QEE CPs are between 94% and 95% (see, for example, Table 2 when $k = 73$). R_2 , G_2 , WA_1 , and WA_2 have reasonable coverage properties, but for small values of π and k , these methods

Figure 4: The expected interval length of 95% nominal confidence intervals for the proportion π based on the 11 methods for fixed litter sizes. Each box plot was constructed based on 10 parameter combinations.

tend to be very conservative, whereas WA_1 and WA_2 tend to be very liberal in a few cases (see, for example, Table 2 for $\phi = 0.5$ and $0.25 \leq \pi \leq 0.35$ when $k = 20$). For small values of k , the CPs for MCP are larger than 96% for almost all situations, while the CPs for MCP are between 95% and 96% for larger values of k . Overall, the WI_1 , WI_2 , PL, and MWS intervals have the best coverage properties, meaning the CPs of these methods are very close to the nominal level in almost all the data situations considered here. However, in most situations, WI_1 , WI_2 , and MWS show better consistency in coverage than PL.

As expected, we see from the results in Figures 4 and 5 and Table 3 that the ELs for all the methods decrease as k increases. In addition, the ELs for all the methods increase when the risk rate, as well as the deviation from independence among observations within the same cluster, increase. For instance,

Figure 5: The expected interval length of 95% nominal confidence intervals for the proportion π based on the 11 methods for ED litter sizes. Each box plot was constructed based on 10 parameter combinations.

ranges for the ELs in Figure 4(v) for $\phi = 0.5$ are larger than those in Figure 4(v) for $\phi = 0.1$, and the ELs for all the methods when $\pi = 0.05$ are much larger than those when $\pi = 0.45$ (see, Table 3). It should be noted that there are no significant differences in the ELs between the fixed and variable cluster sizes. In general, ML and PL provide the lowest ELs, while QEE and MCP show the largest ELs, particularly for small values of k . Although the ELs for ML are among the smallest, in some situations this is at the expense of serious under/over-coverage. The remaining seven methods have very similar ELs throughout the parameter space points considered here except for MWS whose ELs were found to be slightly higher than the ELs of the other six methods. Further details regarding the above results can be found in the Supplementary Materials (see, Tables W1–W4).

Table 2: The coverage probability estimates based on confidence intervals by the methods with nominal level, $1 - \alpha = 95\%$ for fixed and ED litter sizes.

k	ϕ	π	PL	WI ₁	WI ₂	WA ₁	WA ₂	R ₂	G ₂	ML	QEE	MCP	MWS
Fixed litter sizes													
19	0.1	0.05	0.957	0.959	0.960	0.978	0.979	0.978	0.974	0.984	0.986	0.989	0.969
		0.15	0.943	0.965	0.963	0.955	0.953	0.966	0.960	0.948	0.962	0.982	0.970
		0.25	0.932	0.952	0.947	0.945	0.936	0.951	0.944	0.943	0.944	0.970	0.959
		0.35	0.950	0.949	0.939	0.943	0.934	0.945	0.939	0.955	0.956	0.968	0.957
		0.45	0.940	0.953	0.943	0.946	0.938	0.946	0.940	0.947	0.945	0.968	0.958
	0.3	0.05	0.951	0.961	0.967	0.961	0.969	0.989	0.986	0.990	0.992	0.990	0.967
		0.15	0.946	0.962	0.964	0.943	0.939	0.957	0.950	0.922	0.940	0.981	0.966
		0.25	0.941	0.951	0.944	0.934	0.929	0.938	0.931	0.914	0.928	0.970	0.956
		0.35	0.945	0.944	0.940	0.934	0.927	0.932	0.926	0.933	0.940	0.971	0.958
		0.45	0.932	0.948	0.942	0.938	0.932	0.937	0.931	0.941	0.944	0.967	0.956
	0.5	0.05	0.947	0.954	0.964	0.944	0.961	0.990	0.988	0.994	0.996	0.992	0.961
		0.15	0.949	0.958	0.962	0.934	0.933	0.946	0.941	0.916	0.940	0.979	0.961
		0.25	0.944	0.949	0.951	0.925	0.926	0.929	0.923	0.899	0.917	0.975	0.961
		0.35	0.947	0.950	0.948	0.935	0.932	0.936	0.928	0.935	0.948	0.972	0.960
		0.45	0.949	0.949	0.948	0.933	0.933	0.931	0.9267	0.932	0.937	0.972	0.958
73	0.1	0.05	0.937	0.954	0.954	0.944	0.945	0.951	0.950	0.953	0.971	0.964	0.954
		0.15	0.935	0.947	0.942	0.945	0.942	0.943	0.941	0.940	0.951	0.953	0.946
		0.25	0.939	0.949	0.945	0.949	0.943	0.945	0.944	0.941	0.943	0.957	0.951
		0.35	0.941	0.950	0.944	0.948	0.943	0.945	0.943	0.943	0.943	0.955	0.949
		0.45	0.948	0.958	0.952	0.955	0.950	0.952	0.951	0.966	0.968	0.961	0.956
	0.3	0.05	0.934	0.951	0.953	0.935	0.933	0.942	0.941	0.970	0.983	0.962	0.948
		0.15	0.949	0.950	0.951	0.945	0.945	0.945	0.944	0.922	0.941	0.959	0.949
		0.25	0.931	0.953	0.952	0.948	0.946	0.945	0.943	0.934	0.946	0.958	0.950
		0.35	0.944	0.951	0.952	0.949	0.951	0.945	0.943	0.942	0.944	0.959	0.954
		0.45	0.946	0.956	0.953	0.952	0.951	0.948	0.946	0.967	0.967	0.959	0.953
	0.5	0.05	0.951	0.954	0.961	0.945	0.943	0.949	0.947	0.977	0.989	0.971	0.954
		0.15	0.941	0.950	0.952	0.940	0.944	0.939	0.938	0.915	0.936	0.962	0.950
		0.25	0.933	0.952	0.958	0.946	0.950	0.944	0.943	0.933	0.950	0.963	0.954
		0.35	0.939	0.951	0.956	0.946	0.952	0.943	0.942	0.936	0.939	0.960	0.950
		0.45	0.943	0.953	0.957	0.950	0.953	0.946	0.944	0.962	0.965	0.961	0.952
ED litter sizes													
20	0.1	0.05	0.962	0.965	0.967	0.972	0.975	0.984	0.980	0.981	0.986	0.991	0.974
		0.15	0.939	0.955	0.951	0.943	0.941	0.954	0.946	0.940	0.950	0.974	0.962
		0.25	0.936	0.947	0.941	0.940	0.935	0.943	0.938	0.940	0.941	0.967	0.960
		0.35	0.930	0.944	0.938	0.937	0.931	0.941	0.936	0.943	0.944	0.966	0.956
		0.45	0.932	0.946	0.938	0.942	0.934	0.943	0.938	0.951	0.951	0.963	0.955
	0.3	0.05	0.947	0.958	0.967	0.954	0.961	0.986	0.982	0.990	0.995	0.990	0.966
		0.15	0.943	0.959	0.962	0.935	0.934	0.947	0.941	0.905	0.924	0.980	0.966
		0.25	0.949	0.945	0.944	0.934	0.931	0.938	0.932	0.913	0.926	0.968	0.956
		0.35	0.953	0.938	0.938	0.927	0.926	0.929	0.923	0.927	0.933	0.962	0.951
		0.45	0.942	0.944	0.941	0.935	0.934	0.937	0.930	0.944	0.949	0.966	0.954
	0.5	0.05	0.943	0.953	0.963	0.956	0.961	0.982	0.978	0.993	0.997	0.989	0.959
		0.15	0.950	0.957	0.965	0.935	0.936	0.942	0.938	0.897	0.925	0.981	0.966
		0.25	0.951	0.943	0.947	0.921	0.923	0.926	0.921	0.901	0.918	0.970	0.957
		0.35	0.952	0.943	0.944	0.928	0.932	0.931	0.925	0.934	0.948	0.969	0.953
		0.45	0.953	0.941	0.947	0.930	0.934	0.932	0.926	0.938	0.944	0.971	0.957
50	0.1	0.05	0.952	0.952	0.956	0.940	0.942	0.953	0.951	0.960	0.978	0.967	0.956
		0.15	0.938	0.949	0.947	0.945	0.944	0.947	0.945	0.943	0.952	0.959	0.951
		0.25	0.939	0.948	0.945	0.945	0.941	0.944	0.942	0.941	0.943	0.959	0.951
		0.35	0.939	0.946	0.943	0.943	0.939	0.943	0.941	0.941	0.942	0.955	0.948
		0.45	0.943	0.946	0.944	0.945	0.942	0.944	0.941	0.960	0.961	0.957	0.950

(continued)

Table 2: (continued)

k	ϕ	π	PL	WI ₁	WI ₂	WA ₁	WA ₂	R ₂	G ₂	ML	QEE	MCP	MWS
	0.3	0.05	0.954	0.958	0.967	0.933	0.936	0.950	0.949	0.981	0.989	0.973	0.959
		0.15	0.951	0.944	0.948	0.935	0.936	0.939	0.936	0.915	0.930	0.958	0.948
		0.25	0.934	0.952	0.954	0.946	0.946	0.946	0.944	0.933	0.944	0.964	0.956
		0.35	0.941	0.951	0.952	0.947	0.948	0.946	0.944	0.941	0.943	0.961	0.953
		0.45	0.945	0.952	0.954	0.948	0.951	0.946	0.944	0.964	0.966	0.962	0.955
	0.5	0.05	0.950	0.957	0.968	0.937	0.938	0.946	0.945	0.983	0.991	0.977	0.958
		0.15	0.947	0.949	0.955	0.933	0.939	0.937	0.934	0.906	0.930	0.966	0.954
		0.25	0.948	0.949	0.955	0.938	0.945	0.940	0.937	0.927	0.943	0.965	0.954
		0.35	0.938	0.949	0.956	0.940	0.950	0.940	0.937	0.937	0.940	0.962	0.954
		0.45	0.937	0.946	0.952	0.942	0.947	0.942	0.939	0.955	0.959	0.961	0.952

Table 3: The expected interval lengths based on confidence intervals by the methods with nominal level, $1 - \alpha = 95\%$ for fixed and ED litter sizes.

k	ϕ	π	PL	WI ₁	WI ₂	WA ₁	WA ₂	R ₂	G ₂	ML	QEE	MCP	MWS
Fixed litter sizes													
19	0.1	0.05	0.105	0.113	0.111	0.105	0.104	0.104	0.101	0.122	0.211	0.128	0.119
		0.15	0.150	0.164	0.158	0.165	0.158	0.161	0.157	0.159	0.253	0.193	0.178
		0.25	0.181	0.195	0.186	0.199	0.190	0.194	0.188	0.191	0.213	0.229	0.212
		0.35	0.199	0.213	0.203	0.218	0.208	0.213	0.207	0.209	0.210	0.251	0.233
		0.45	0.209	0.222	0.211	0.228	0.216	0.222	0.216	0.224	0.225	0.261	0.243
	0.3	0.05	0.133	0.147	0.144	0.134	0.131	0.131	0.127	0.148	0.183	0.163	0.151
		0.15	0.183	0.213	0.204	0.214	0.205	0.208	0.203	0.191	0.327	0.242	0.220
		0.25	0.227	0.251	0.240	0.259	0.247	0.252	0.245	0.235	0.255	0.285	0.259
		0.35	0.256	0.275	0.262	0.286	0.272	0.278	0.271	0.261	0.269	0.312	0.284
		0.45	0.270	0.287	0.273	0.300	0.285	0.291	0.283	0.275	0.278	0.325	0.296
	0.5	0.05	0.166	0.179	0.175	0.163	0.157	0.160	0.156	0.172	0.205	0.199	0.182
		0.15	0.214	0.253	0.244	0.255	0.246	0.250	0.243	0.222	0.298	0.286	0.257
		0.25	0.264	0.296	0.286	0.309	0.297	0.303	0.295	0.273	0.317	0.336	0.301
		0.35	0.299	0.324	0.312	0.342	0.329	0.336	0.327	0.306	0.318	0.367	0.329
		0.45	0.316	0.336	0.325	0.358	0.344	0.351	0.342	0.321	0.329	0.381	0.342
73	0.1	0.05	0.040	0.045	0.044	0.044	0.043	0.043	0.043	0.049	0.143	0.048	0.045
		0.15	0.067	0.072	0.070	0.072	0.070	0.071	0.070	0.069	0.076	0.076	0.074
		0.25	0.083	0.087	0.085	0.088	0.085	0.086	0.085	0.084	0.085	0.092	0.089
		0.35	0.093	0.096	0.094	0.097	0.094	0.095	0.094	0.093	0.093	0.101	0.098
		0.45	0.097	0.100	0.098	0.101	0.098	0.099	0.098	0.099	0.100	0.106	0.103
	0.3	0.05	0.051	0.064	0.063	0.062	0.061	0.061	0.060	0.061	0.070	0.067	0.063
		0.15	0.088	0.103	0.100	0.103	0.100	0.100	0.100	0.091	0.362	0.106	0.101
		0.25	0.112	0.124	0.120	0.125	0.121	0.122	0.121	0.115	0.121	0.128	0.122
		0.35	0.128	0.137	0.133	0.138	0.134	0.135	0.134	0.129	0.132	0.140	0.135
		0.45	0.135	0.142	0.138	0.144	0.140	0.141	0.140	0.136	0.137	0.146	0.141
	0.5	0.05	0.063	0.080	0.078	0.077	0.075	0.076	0.075	0.072	0.085	0.084	0.079
		0.15	0.105	0.125	0.122	0.125	0.122	0.123	0.122	0.108	0.144	0.130	0.123
		0.25	0.134	0.151	0.148	0.152	0.149	0.150	0.149	0.137	0.154	0.156	0.148
		0.35	0.152	0.166	0.162	0.168	0.164	0.165	0.164	0.155	0.161	0.171	0.163
		0.45	0.162	0.173	0.169	0.176	0.172	0.173	0.172	0.163	0.168	0.178	0.170
ED litter sizes													
20	0.1	0.05	0.084	0.090	0.090	0.086	0.086	0.085	0.083	0.101	0.190	0.103	0.097
		0.15	0.106	0.137	0.133	0.137	0.134	0.136	0.132	0.133	0.157	0.157	0.147
		0.25	0.143	0.164	0.159	0.166	0.161	0.164	0.160	0.162	0.164	0.188	0.176

(continued)

Table 3: (continued)

k	ϕ	π	PL	WI ₁	WI ₂	WA ₁	WA ₂	R ₂	G ₂	ML	QEE	MCP	MWS
50	0.3	0.35	0.165	0.180	0.174	0.183	0.177	0.181	0.177	0.180	0.181	0.206	0.194
		0.45	0.178	0.187	0.181	0.191	0.184	0.189	0.184	0.194	0.195	0.214	0.201
		0.05	0.114	0.128	0.128	0.118	0.118	0.118	0.115	0.127	0.147	0.145	0.134
		0.15	0.164	0.192	0.187	0.193	0.188	0.192	0.187	0.172	0.247	0.218	0.200
		0.25	0.208	0.228	0.222	0.234	0.227	0.232	0.226	0.214	0.232	0.259	0.237
		0.35	0.235	0.249	0.242	0.257	0.249	0.255	0.249	0.240	0.245	0.282	0.259
	0.5	0.45	0.249	0.260	0.252	0.269	0.261	0.268	0.261	0.252	0.256	0.294	0.271
		0.05	0.147	0.164	0.162	0.151	0.147	0.151	0.147	0.153	0.182	0.188	0.172
		0.15	0.195	0.234	0.229	0.236	0.230	0.236	0.230	0.204	0.278	0.270	0.244
		0.25	0.246	0.274	0.269	0.284	0.278	0.284	0.277	0.255	0.306	0.315	0.285
		0.35	0.280	0.300	0.294	0.314	0.308	0.315	0.307	0.287	0.303	0.344	0.311
		0.45	0.297	0.312	0.306	0.329	0.322	0.329	0.321	0.302	0.313	0.358	0.323
	0.1	0.05	0.049	0.054	0.054	0.053	0.052	0.052	0.052	0.060	0.079	0.059	0.056
		0.15	0.081	0.086	0.085	0.086	0.085	0.085	0.085	0.084	0.110	0.093	0.090
		0.25	0.100	0.104	0.103	0.105	0.103	0.104	0.103	0.102	0.103	0.113	0.109
		0.35	0.111	0.115	0.112	0.115	0.113	0.114	0.113	0.113	0.113	0.124	0.119
		0.45	0.117	0.120	0.117	0.121	0.118	0.119	0.118	0.121	0.121	0.129	0.124
		0.05	0.063	0.077	0.077	0.074	0.073	0.073	0.073	0.075	0.087	0.083	0.078
	0.3	0.15	0.106	0.122	0.120	0.122	0.120	0.121	0.120	0.110	0.207	0.130	0.123
		0.25	0.135	0.148	0.145	0.149	0.147	0.148	0.146	0.139	0.145	0.156	0.148
		0.35	0.154	0.163	0.160	0.165	0.162	0.163	0.162	0.156	0.159	0.172	0.163
		0.45	0.162	0.170	0.167	0.172	0.169	0.170	0.169	0.163	0.166	0.179	0.170
		0.05	0.079	0.098	0.096	0.093	0.091	0.092	0.091	0.089	0.105	0.105	0.098
		0.15	0.126	0.150	0.147	0.150	0.148	0.149	0.148	0.130	0.177	0.159	0.149
	0.5	0.25	0.161	0.180	0.177	0.183	0.180	0.182	0.180	0.165	0.188	0.191	0.179
		0.35	0.183	0.198	0.195	0.202	0.199	0.201	0.199	0.186	0.194	0.210	0.197
		0.45	0.194	0.206	0.203	0.211	0.208	0.209	0.207	0.196	0.202	0.218	0.205

5 Applications to real data analysis

We apply all 15 CI methods considered in Section 3 to three data sets: (i) CTC images data [8], (ii) chronic obstructive pulmonary disease (COPD) data [2], and (iii) screening mammogram data [9]. The number of clusters and cluster sizes varied among these studies.

5.1 Example 1: CTC images data

This study was a contrast-enhanced multi-detector row spiral computed tomography coronary (CTC) angiography [8], which is an imaging test that can detect polyps before they develop into cancer. Investigators developed a computer algorithm, called computer aided detection (CAD), to help radiologists detect polyps on the CTC. The main purpose of this study was to assess the radiologists' diagnostic accuracy of CAD-enhanced CTC for detecting polyps. In the trial, 270 patients from six institutions were compiled in the retrospective design. These patients had undergone CTC for several medical reasons. In order to assess the radiologists' performance, 25 patients were randomly selected from the 119 test cases. This study showed that there were actually multiple polyps in some of those patients varying, from 1 to 3 polyps with sample mean 1.56 and standard deviation 0.58, which indicates that the detection capabilities for each patient may be correlated. In order to assess the radiologists' diagnostic accuracy in this study, we considered the confidence interval procedures discussed in Section 3 to estimate the sensitivity of CAD-enhanced CTC for detecting polyps.

The values (standard error) of $\hat{\pi}_{ml}$, $\hat{\pi}_{eq}$, $\hat{\pi}_{de}$, and $\hat{\pi}_{qee}$ are given by 0.8464 (0.0633), 0.8473 (0.0614), 0.8456 (0.0777), and 0.8433 (0.0685), respectively, and the values of $\hat{\phi}_{ml}$ and $\hat{\phi}_a$ are 0.3426 and 0.4885, respectively. Note that the estimated sensitivity of CAD-enhanced CTC for detecting polyps is very similar among these four methods, whereas the estimated intraclass correlation from ML is lower than for ANOVA. The 95% asymptotic CIs for the true sensitivity of CAD-enhanced CTC for detecting polyps are provided in Table 4. From the results, we see that all 15 CIs for sensitivity of CAD-enhanced CTC for detecting polyps lead to similar conclusions. It may be noted, however, that the CIs for sensitivity based on PL, WI_2 , R_2 , and G_2 are very similar and slightly shorter than the other intervals. Moreover, the lengths of the PL and ML intervals are similar and shortest, which is in close agreement with the simulation results presented earlier.

Table 4: Ninety-five percent confidence intervals of the single proportion π by all 15 methods for the data in Example 1.

Method	Lower CI	Upper CI	Length	Length comparison individual/ML
PL	0.6982	0.9419	0.2437	0.982
WI_1	0.6735	0.9362	0.2626	1.058
WI_2	0.6821	0.9340	0.2519	1.015
WA_1	0.7133	0.9790	0.2656	1.070
WA_2	0.7192	0.9736	0.2544	1.025
R_1	0.7093	0.9657	0.2564	1.033
R_2	0.6853	0.9426	0.2574	1.037
G_1	0.7119	0.9631	0.2512	1.012
G_2	0.6879	0.9400	0.2522	1.016
ML	0.7223	0.9705	0.2482	1.000
EQL	0.7269	0.9676	0.2407	0.970
DEQL	0.6933	0.9978	0.3046	1.227
QEE	0.7090	0.9776	0.2686	1.082
MCP	0.6571	0.9551	0.2980	1.2007
MWS	0.6697	0.9376	0.2679	1.0794

5.2 Example 2: Chronic obstructive pulmonary disease data

This was a case-control study of familial aggregation of chronic obstructive pulmonary disease (COPD). The main object of this study was to determine whether there is a significant probability that a given sibling of a COPD patient has impaired pulmonary function (IPF). The data given in Table 5 of Liang *et al.* ([2]) refer to the frequency distribution of the number of IPF cases per family. There are 203 siblings from 100 families of various sizes ranging from 1 to 6 with the sample mean 2.03 and standard deviation 1.32. The binary response of interest is whether a given sibling of a COPD patient has IPF. Siblings within each family are correlated and the number of IPF cases per family may be over-dispersed as compared to a binomial model. To examine the existence of IPF for a given sibling of this study, we considered the confidence interval procedures discussed in Section 3 to estimate the IPF rate.

We compute the values (standard error) of $\hat{\pi}_{ml}$, $\hat{\pi}_{eq}$, $\hat{\pi}_{de}$, and $\hat{\pi}_{qee}$ as 0.2822 (0.0362), 0.2751 (0.0390), 0.2829 (0.0433), and 0.2854 (0.0365), respectively. The ML and ANOVA estimates of ϕ are given by $\hat{\phi}_{ml} = 0.2128$ and $\hat{\phi}_a = 0.1871$. We also computed 95% asymptotic CIs for the IPF rate π using all 15 methods considered here, which are presented in Table 5. From the results in this table, it is evident that all the intervals tend to support the existence of IPF for a given sibling of a COPD patient. Note that the PL, WI_1 , WI_2 , WA_1 , WA_2 , ML, QEE, and MWS interval lengths are smaller than those of the other methods; however, the PL and ML intervals have slightly shorter lengths.

Table 5: Ninety-five percent confidence intervals of the single proportion π by all 15 methods for the data in Example 2.

Method	Lower CI	Upper CI	Length	Length comparison individual/ML
PL	0.2150	0.3563	0.1413	0.995
WI ₁	0.2284	0.3729	0.1445	1.018
WI ₂	0.2153	0.3603	0.1451	1.021
WA ₁	0.2226	0.3686	0.1460	1.028
WA ₂	0.2089	0.3555	0.1465	1.032
R ₁	0.2174	0.3758	0.1584	1.115
R ₂	0.2204	0.3786	0.1582	1.114
G ₁	0.2178	0.3754	0.1576	1.110
G ₂	0.2208	0.3782	0.1574	1.109
ML	0.2112	0.3532	0.1420	1.000
EQL	0.1987	0.3514	0.1527	1.075
DEQL	0.1979	0.3678	0.1699	1.196
QEE	0.2139	0.3570	0.1431	1.007
MCP	0.1938	0.3484	0.1546	1.089
MWS	0.1988	0.3458	0.1470	1.035

A third application of the methods to the analysis of a screening mammogram dataset is given in the Supplementary Materials.

6 Discussion and concluding remarks

This paper considers 13 asymptotic CIs for binary outcome data, taken from clusters, assuming a beta-binomial distribution and semi-parametric models such that only the first two moments of the responses need be specified. The results of our simulation studies in Section 4 suggest that the two versions of the Wilson score, the modified Wilson score, and the profile likelihood methods are preferable as their observed CPs are very close to the nominal coverage level. However, the Wilson score and the modified Wilson score methods are preferred to the profile likelihood method in that they are well controlled around the desired coverage level, though the profile likelihood method is preferred to the two versions of the Wilson score and the modified Wilson score in the sense that it generally possesses shorter ELs in almost all data situations.

Our results in this paper depend on the assumption of the beta-binomial distribution and semi-parametric models specified by only the first two moments of the response variable (which we considered similar to the mean and variance of the beta-binomial distribution). Moreover, we considered a common correlation structure assumption among the observations within the same cluster which is often applicable to family studies; however, one can extend this research by using a more general correlation structure arising in many genetic epidemiology studies.

In our simulation studies it appeared that in some situations the asymptotic confidence intervals based on ML, EQL, DEQL, and QEE showed serious lack of coverage, particularly EQL and DEQL. These CI procedures rely largely on the asymptotic normality distribution assumption for the estimate of the parameter, which may not hold in some situations, especially for small sample sizes or small parameter values. Moreover, the problem of lack of coverage could arise due to estimation of the asymptotic standard error. Further research that can solve these issues by incorporating alternative distributional approximations, such as parametric, nonparametric, and double bootstrap, to yield greater coverage accuracy would clearly be worthwhile.

Acknowledgments: The authors are grateful to the Associate Editor and two referees for their constructive comments and suggestions that have led to much improvement in this manuscript. This research was supported in part by a CSU-AAUP University research grant.

References

1. Paul SR. Analysis of proportions of affected fetuses in teratological experiments. *Biometrics* 1982;38:361–70.
2. Liang KY, Qaqish B, Zeger SL. Multivariate regression analysis for categorical data. *J Roy Stat Soc B* 1992;54:3–40.
3. Saha KK. Profile likelihood-based confidence intervals of the intraclass correlation for binary outcome data sampled from clusters. *Stat Med* 2012;31:3982–4002.
4. Hogg DR, Tanis DV. Introduction to probability statistics. London: Chapman and Hall, 2009.
5. Rao JNK, Scott AJ. A simple method for the analysis of clustered binary data. *Biometrics* 1992;48:577–85.
6. Kleinman J. Proportions with extraneous variance: single and independent sample. *J Am Stat Assoc* 1973;68:46–54.
7. Paul SR, Islam AS. Joint estimation of the mean and dispersion parameters in the analysis of proportions: a comparison of efficiency and bias. *Can J Stat* 1998;26:83–94.
8. Zhou X-H, Obuchowski NA, McClish DK. Statistical methods in diagnostic medicine. New York: Wiley, 2002.
9. Kim J, Lee JH. Simultaneous confidence intervals for a success probability and intraclass correlation, with an application to screening mammography. *Biometric J* 2013;55:944–54.
10. Agresti A, Coull BA. Approximate is better than ‘exact’ for interval estimation of binomial proportions. *Am Statistician* 1998;52:119–26.
11. Newcombe RG. Two sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 1998;17:857–72.
12. Korn EL, Graubard BI. Analysis of health surveys. New York: John Wiley & Sons, 1999.
13. Lui KJ. Statistical estimation of epidemiological risk. New York: Wiley, 2004.
14. Rutter CM. Bootstrap estimation of diagnostic accuracy with patient-clustered data. *Acad Radiol* 2000;2:413–19.
15. Paul SR, Zaihra T. Interval estimation of risk difference for data sampled from clusters. *Stat Med* 2008;27:4207–20.
16. Williams DA. Analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics* 1975;31:949–52.
17. Paul SR, Islam AS. Analysis of proportions in the presence of over-/under-dispersion. *Biometrics* 1995;51:1400–11.
18. Liang KY, McCullagh P. Case studies in binary dispersion. *Biometrics* 1993;49:623–30.
19. Altaye M, Donner A, Klar N. Inference procedures for assessing interobserver agreement among multiple raters. *Biometrics* 2001;57:584–8.
20. Pradhan V, Saha KK, Banerjee T, Evans J. Weighted profile likelihood-based confidence interval for the difference between two proportions with paired binomial data. *Stat Med* 2014;33:2984–97.
21. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986;42:121–30.
22. Paul SR, Saha KK. The generalized linear model and extensions: a review and some biological and environmental applications. *Environmetrics* 2007;18:421–43.
23. Inagaki N. Asymptotic relations between the likelihood estimating function and the maximum likelihood estimator. *Ann Inst Stat Math* 1973;25:1–26.
24. Bowman D, George EO. A saturated model for analyzing exchangeable binary data: Applications to clinical and developmental toxicity studies. *J Am Stat Assoc* 1995;90:871–9.
25. Kupper LL, Portier C, Hogan MD, Yamamoto E. The impact of litter effects on dose-response modeling in teratology. *Biometrics* 1986;42:85–98.