Available online at www.sciencedirect.com
ScienceDirect
Contemporary Clinical Trials 28 (2007) 120–129
Contemporary Clinical Trials
www.elsevier.com/locate/conclintrial

Interval estimation of risk ratio in the simple compliance randomized trial

Kung-Jong Lui *

Department of Mathematics and Statistics, San Diego State University, San Diego, CA 92182-7720, USA

Received 5 July 2005; accepted 22 May 2006

Abstract

Consider the simple compliance randomized trial (SCRT), in which patients assigned to an experimental group may switch to receive a control treatment, but patients assigned to a control group are assumed to all receive their assigned treatment. We develop five asymptotic interval estimators for the relative risk (RR) of probabilities of response among patients who would comply with the experimental treatment under the SCRT. We employ Monte Carlo simulation to evaluate the performance of these interval estimators in a variety of situations. We note that the interval estimator using Wald's statistic and the interval estimator derived from a quadratic equation based on asymptotic properties of the maximum likelihood estimator (MLE) can lose accuracy, while the most commonly-used interval estimator using a logarithmic transformation of the MLE for the RR suggested elsewhere can lose efficiency. We further note that the probability of failure to apply the interval estimator derived from an idea used in Fieller's Theorem to produce a confidence interval can be non-negligible even when the number of patients in both comparison groups is not small. Finally, we find that an interval estimator using a simple ad hoc procedure of combining two interval estimators with and without a logarithmic transformation of the MLE can consistently perform well with respect to the coverage probability even when the number of patients per treatment is not large. In fact, this estimator uniformly outperforms all the other estimators considered here and thereby is recommended for general use. We include an example regarding the study of vitamin A supplementation to reduce the mortality among preschool children to illustrate the use of interval estimators discussed in this paper.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Efficiency; Simple compliance study; Coverage probability; Efficacy; Relative risk

1. Introduction

Consider the simple compliance randomized trial (SCRT), in which patients assigned to an experimental group may switch to receive a control treatment (or placebo), while patients assigned to a control group (or placebo) are assumed to all receive their assigned treatment. For example, consider the randomized trial studying vitamin A supplementation to reduce mortality among preschool children in rural Indonesia [1]. Children were randomly assigned to either the experimental group of receiving a large oral dose of vitamin A 2 to 3 months following baseline enumeration and again 6 months later, or to a control group of receiving no vitamin A supplementation. Nearly 20% of children assigned to the

* Tel.: +1 619 594 6191; fax: +1 619 594 7239.
E-mail address: kjl@rohan.sdsu.edu.

1551-7144/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.cct.2006.05.005

experimental group failed to receive vitamin A as prescribed (and hence these children could be regarded as receiving the same treatment as that for children in the control group). On the other hand, children assigned to the control group would receive, by design, no vitamin A supplementation. Deaths were ascertained in a second population census 12 months following the baseline census. Mortality rates in children between these two comparison groups were then compared. Directly using the ratio of mortality rates in the SCRT based on an intent-to-treat (ITT) analysis to estimate the biological efficacy of vitamin A supplementation in reducing mortality is biased [2]. Although we may find an inspiring discussion [2] on estimation of the biological efficacy, the discussion on interval estimation of the relative risk (RR) of mortality rates in such a SCRT is limited.

In this paper, we develop five asymptotic interval estimators for the RR in a SCRT. These include the interval estimator using Wald's statistic [3], the interval estimator using a logarithmic transformation of the maximum likelihood estimator (MLE) [2], the interval estimator using an idea similar to that in Fieller's Theorem [3], the interval estimator calculated by use of a quadratic equation derived from asymptotic properties of the MLE, and the interval estimator using a simple ad hoc procedure of combining two interval estimators with and without using a logarithmic transformation. We apply Monte Carlo simulation to evaluate the finite sample performance of these five interval estimators with respect to the coverage probability and average length in a variety of situations. Finally, we use the example [1,2] regarding the study of vitamin A to reduce the mortality rate in preschool children to illustrate the practical use of these interval estimators. Note that because the data structure for the single consent randomized trial [4–13] is essentially parallel to that for the SCRT, all interval estimators for the RR discussed here are applicable in the former as well.

2. Notations and interval estimators

Consider comparing an experimental treatment with a control treatment in a SCRT. Suppose that we randomly assign patients to receive one of these two treatments under comparison. Patients assigned to the experimental treatment may decline the assigned therapy and switch to receive the control treatment. On the other hand, patients assigned to the control treatment are assumed to all receive their assigned treatment. For clarity, we summarize in the following 2×2 table the cell probability $p_{ij}^{(T)}$, where the superscript T denotes the treatment status: $T=E$ and $T=C$ for the experimental and control treatments, respectively. For example, we let $p_{11}^{(T)}$ denote the probability of obtaining a patient who would comply with the experimental treatment and have the outcome of interest if he/she were assigned to receive treatment T. In particular, $p_{11}^{(C)}$ here denotes in the control group the probability of obtaining a patient who would comply with the experimental treatment (if he/she were actually assigned to the experimental group) and have a positive response rather than the probability of obtaining a patient who complies with the control assignment and has a positive response. We further let $p_{i+}^{(T)}=p_{i1}^{(I)}+p_{i0}^{(T)}$ and $p_{+j}^{(T)}=p_{1j}^{(T)}+p_{0j}^{(T)}$ and $j=1, 0$, for $T=E,C$.

		Treatment T ($T=E,C$)		
		Compliance with the experimental treatment		
		Yes	No	Total
Response	Yes	$p_{11}^{(T)}$	$p_{10}^{(T)}$	$p_{1+}^{(T)}$
	No	$p_{01}^{(T)}$	$p_{00}^{(T)}$	$p_{0+}^{(T)}$
	Total	$p_{+1}^{(T)}$	$p_{+0}^{(T)}$	1

Because if a patient assigned to the experimental group does not comply with his/her assigned treatment in a SCRT, he/she will receive the control treatment, and because the assignment of patients to either of the two comparison groups is random, we may reasonably assume that $p_{i0}^{(E)}=p_{i0}^{(C)}$ for $i=1,0$ [2,6]. These imply that the proportions of patients who would comply with the experimental treatment between two comparison groups are equal (i.e., $p_{+1}^{(E)}=p_{+1}^{(C)}$). We let θ denote this common proportion. Suppose that we independently assign n_E and n_C patients to the experimental and control groups, respectively. Let $n_{ij}^{(T)}$ denote the observed frequency corresponding to the cell probability $p_{ij}^{(T)}$ in treatment T(= E or C). The random vector $\underline{n}_E=(n_{11}^{(E)}, n_{10}^{(E)}, n_{01}^{(E)}, n_{00}^{(E)})'$ then follows the multinomial distribution with parameters n_E and $(p_{11}^{(E)}, p_{10}^{(E)}, p_{01}^{(E)}, p_{00}^{(E)})'$. Because we do not have the information on patients assigned to the control group about whether they would comply with the experimental treatment if they were actually assigned to receive this therapy, we can only observe the marginal total number of responses $n_{1+}^{(C)}(= n_{11}^{(C)}+n_{10}^{(C)})$, which follows the binomial distribution with parameters n_C and $p_{1+}^{(C)}$ $(= p_{11}^{(C)}+p_{10}^{(C)})$. We can easily see that the MLEs for $p_{ij}^{(C)}$ and $p_{1+}^{(C)}$ are given by $\hat{p}_{ij}^{(C)}=n_{ij}^{(C)}/n_E$, and $\hat{p}_{1+}^{(C)}=n_{1+}^{(C)}/n_C$, respectively. By the functional invariance property [3], the MLE for $\theta(= p_{+1}^{(C)}=p_{+1}^{(C)})$

is simply $\hat{\theta}=\hat{p}_{+1}^{(C)}$, where $\hat{p}_{+1}^{(C)}=\hat{p}_{11}^{(C)}+\hat{p}_{01}^{(C)}$. In this paper, following Sommer and Zeger [2], we are interested in interval estimation on the RR of probabilities of response $\gamma=(p_{11}^{(C)}/p_{+1}^{(C)})/(p_{11}^{(C)}/p_{+1}^{(C)})=p_{11}^{(C)}/p_{11}^{(C)}$ among patients who would comply with the experimental treatment. The range for γ is $0<\gamma<\infty$.

First, note that the MLE of the RR is given by [2]

$$\hat{\gamma}=\hat{p}_{11}^{(E)}/(\hat{p}_{1+}^{(C)}-\hat{p}_{10}^{(E)}). \tag{1}$$

When applying the delta method, we can show that the asymptotic variance of $\hat{\gamma}$ is

$$Var(\hat{\gamma})=\gamma^2\{(1-p_{11}^{(E)})/(n_E p_{11}^{(E)})+[p_{1+}^{(C)}(1-p_{1+}^{(C)})/n_C+p_{10}^{(E)}(1-p_{10}^{(E)})/n_E]/(p_{1+}^{(C)}-p_{10}^{(E)})^2-2p_{10}^{(E)}/(n_E(p_{1+}^{(C)}-p_{10}^{(E)}))\}. \tag{2}$$

On the basis of Wald's statistic, we obtain an asymptotic $100(1-\alpha)$ percent confidence interval for γ as

$$[\max\{\hat{\gamma}-Z_{\alpha/2}(\hat{Var}(\hat{\gamma}))^{1/2},0\},\quad \hat{\gamma}+Z_{\alpha/2}(\hat{Var}(\hat{\gamma}))^{1/2}], \tag{3}$$

where $\max\{a,b\}$ denotes the maximum of a and b, and $\hat{\mathrm{V}}\mathrm{ar}(\hat{\gamma})$ is obtained by substituting the MLEs $\hat{\gamma}$ for γ, $\hat{p}_{1+}^{(C)}$ for $p_{1+}^{(C)}$, and $\hat{p}_{ij}^{(E)}$ for $p_{ij}^{(E)}$ in $Var(\hat{\gamma})$(2). To improve the normal approximation of the sampling distribution for $\hat{\gamma}$, we commonly consider use of a logarithmic transformation [2,14]. This leads us to obtain an asymptotic $100(1-\alpha)$ percent confidence interval for γ as [2]

$$[\hat{\gamma}\exp(-Z_{\alpha/2}(\hat{Var}(\log(\hat{\gamma})))^{1/2}),\quad \hat{\gamma}\exp(Z_{\alpha/2}(\hat{Var}(\log(\hat{\gamma})))^{1/2})], \tag{4}$$

where $\hat{\mathrm{V}}\mathrm{ar}(\log(\hat{\gamma}))=\hat{\mathrm{V}}\mathrm{ar}(\hat{\gamma})/\hat{\gamma}^2$.

Note that the MLE $\hat{\gamma}$ (Eq. (1)) is a ratio of two random variables and hence the sampling distribution of $\hat{\gamma}$ can be skewed when both n_E and n_C are not large. To avoid this concern of skewness in normal approximation on which interval estimator (Eq. (3)) depends, we consider $Z(\gamma)=\hat{p}_{11}^{(E)}-\gamma(\hat{p}_{1+}^{(C)}-\hat{p}_{10}^{(E)})$ based on an idea used in Fieller's Theorem [3]. Since $Z(\gamma)$ is a difference (rather than a ratio) of two random variables, the sampling distribution of $Z(\gamma)$ is likely to be less skewed than the sampling distribution of $\hat{\gamma}$ (Eq. (1)). We can easily show that the expectation $E(Z(\gamma))=0$. Furthermore, when using the delta method, we can derive the following asymptotic variance of $Z(\gamma)$ to be

$$Var(Z(\gamma))=p_{11}^{(E)}(1-p_{11}^{(E)})/n_E+\gamma^2[p_{1+}^{(C)}(1-p_{1+}^{(C)})/n_C+p_{10}^{(E)}(1-p_{10}^{(E)})/n_E]-2\gamma p_{11}^{(E)}p_{10}^{(E)}/n_E. \tag{5}$$

As both n_T ($T=E$ or C) are large, we have the probability $P((Z(\gamma))^2/Var(Z(\gamma))\le Z_{\alpha/2}^2)\approx 1-\alpha$. Therefore, using the theorem on page 407 [3], we can claim that the set of γ such that $(Z(\gamma))^2/Var(Z(\gamma))\le Z_{\alpha/2}^2$ forms a $1-\alpha$ test-based confidence region for RR. The above inequality is actually equivalent to the following quadratic equation in γ:

$$A\gamma^2-2B\gamma+C\le 0, \tag{6}$$

where

$$A=(\hat{p}_{1+}^{(C)}-\hat{p}_{10}^{(E)})^2-Z_{\alpha/2}^2[\hat{p}_{1+}^{(C)}(1-\hat{p}_{1+}^{(C)})/n_C+\hat{p}_{10}^{(E)}(1-\hat{p}_{10}^{(E)})/n_E]$$
$$B=\hat{p}_{11}^{(E)}(\hat{p}_{1+}^{(C)}-\hat{p}_{10}^{(E)})-Z_{\alpha/2}^2\hat{p}_{11}^{(E)}\hat{p}_{10}^{(E)}/n_E$$
$$C=(\hat{p}_{11}^{(E)})^2-Z_{\alpha/2}^2\hat{p}_{11}^{(E)}(1-\hat{p}_{11}^{(E)})/n_E.$$

Thus, if $A>0$ and $B^2-AC>0$, then an asymptotic $100(1-\alpha)$ percent confidence interval for γ would be given by

$$[\max\{(B-\sqrt{B^2-AC})/A,0\},(B+\sqrt{B^2-AC})/A]. \tag{7}$$

Note that when $B^2-AC\le 0$, the two distinct real roots of $A\gamma^2-2B\gamma+C=0$ do not exist. Note further that when $A<0$, this quadratic equation is concave and hence the confidence region in this case is a union of two disjoint open intervals (instead of an interval). When either of these cases occurs, we cannot apply Eq. (7) to produce a $100(1-\alpha)$ percent confidence interval for the RR.

Furthermore, on the basis of the asymptotic properties of the MLE $\hat{\gamma}$ (Eq. (1)), we have the probability $P((\hat{\gamma}-\gamma)^2/Var(\hat{\gamma}) \le Z^2_{\alpha/2}) \approx 1-\alpha$ as both n_T are large. The inequality: $(\hat{\gamma}-\gamma)^2/Var(\hat{\gamma}) \le Z^2_{\alpha/2}$ together with $Var(\hat{\gamma})$ Eq. (2) leads us to consider the following quadratic inequality:

$$A^*\gamma^2 - 2B^*\gamma + C^* \le 0, \tag{8}$$

where

$$A^* = 1 + 2Z^2_{\alpha/2}\hat{p}^{(E)}_{10}/[n_E(\hat{p}^{(C)}_{1+}-\hat{p}^{(E)}_{10})]$$
$$B^* = \hat{\gamma}$$
$$C^* = \hat{\gamma}^2\{1-Z^2_{\alpha/2}[(1-\hat{p}^{(E)}_{11})/(n_E\hat{p}^{(E)}_{11}) + (\hat{p}^{(C)}_{1+}(1-\hat{p}^{(C)}_{1+})/n_C + \hat{p}^{(E)}_{10}(1-\hat{p}^{(E)}_{10})/n_E)/(\hat{p}^{(C)}_{1+}-\hat{p}^{(E)}_{10})^2]\}.$$

Note that since the above coefficient $A^*>0$, the quadratic equation $A^*\gamma^2 - 2B^*\gamma + C^* = 0$ is always convex. Furthermore, we can show that $B^{*2} - A^*C^* > 0$ as long as $\hat{\text{Var}}(\hat{\gamma}) > 0$. This suggests that if the MLE $\hat{\gamma}$ exists, the two distinct real roots of the above quadratic equation will exist as well. Thus, an asymptotic $100(1-\alpha)$ percent confidence interval for γ is given by

$$[\max\{(B^* - \sqrt{B^{*2}-A^*C^*})/A^*, 0\},\ (B^* + \sqrt{B^{*2}-A^*C^*})/A^*]. \tag{9}$$

When evaluating the performance of an interval estimator, we commonly use the coverage probability to measure accuracy and the average length to measure efficiency. The coverage probability for an interval estimator is defined as the probability that the random interval $[L(\text{X}), U(\text{X})]$ covers the underlying true parameter, where $L(\text{X})$ and $U(\text{X})$ denote the lower and upper limits, respectively, and are functions of the random vector X of observations in data. In our case, the random vector X is simply equal to $(\underline{\text{n}}'_E, n^{(C)}_{1+})'$. Furthermore, the average length is simply defined as the expectation $E(U(\text{X}) - L(\text{X}))$ for the length of the random interval, where the expectation E is taken with respect to the distribution of X. When an interval estimator has the average length shorter than the other, we say that the former is more efficient than the latter. Note that an interval estimator with a short length would be completely useless if this interval estimator had a very low coverage probability. Similarly, a confidence interval, which has a coverage probability equal to 100% but has a quite wide range, is also practically useless. Note also that the confidence limits, such as using the logarithmic transformation in Eq. (4) or calculating from a quadratic equation in Eq. (7) or Eq. (9), can be un-symmetric with respect to its middle point. Thus, it is possible that the coverage probability of an interval is higher than that of the other, but the former also has the average length shorter than the latter. Therefore, when comparing the performance between interval estimators, we need to simultaneously consider the coverage probability and the average length to avoid drawing a possibly misleading inference exclusively based on the coverage probability [15]. A desirable interval estimator is the interval estimator which has the shortest average length among all interval estimators with the coverage probability larger than or equal to the desired $100(1-\alpha)$ percent confidence level. Although applying the logarithmic transformation of the MLE can often improve the coverage probability of an interval estimator for a ratio of two parameters as for RR, using the logarithmic transformation may cause a loss of efficiency as well. For example, consider the situation in which the denominator $p^{(C)}_{1+} - p^{(E)}_{10}$ of γ is small. The MLE $\hat{\gamma}$, if it exists, can be large by chance, and so is the estimated asymptotic variance $\hat{\text{Var}}(\log(\hat{\gamma}))$. This will produce a quite wide confidence interval using Eq. (4) especially because of involving the exponential transformation. To alleviate the possible loss of accuracy when using Eq. (3) and the possible loss of efficiency when using Eq. (4), we may consider an ad hoc procedure of combining Eqs. (3) and (4). We define the indicator random variable $\delta(\underline{\text{n}}_E, n^{(C)}_{1+})$ to be 1 if the length $\hat{\gamma}[\exp(Z_{\alpha/2}(\hat{\text{Var}}(\log(\hat{\gamma})))^{1/2}) - \exp(-Z_{\alpha/2}(\hat{\text{Var}}(\log(\hat{\gamma})))^{1/2}$ of Eq. (4) is larger than or equal to K times of the length $(\hat{\gamma} + Z_{\alpha/2}\sqrt{\hat{Var}(\hat{\gamma})} - \max\{\hat{\gamma} - Z_{\alpha/2}\sqrt{\hat{Var}(\hat{\gamma})}, 0\})$ of Eq. (3), and 0, otherwise. Thus, we obtain an asymptotic $100(1-\alpha)$ percent confidence interval for γ as

$$[\gamma_l, \gamma_u] \tag{10}$$

where

$$\gamma_l = \delta(\underline{\text{n}}_E, \text{n}^{(\text{C})}_{1+})\max\{\hat{\gamma} - \text{Z}_{\text{a}/2}(\hat{\text{Var}}(\hat{\gamma}))^{1/2}, 0\}$$
$$+ (1-\delta(\underline{\text{n}}_E, \text{n}^{(\text{C})}_{1+}))\hat{\gamma}\exp(-\text{Z}_{\text{a}/2}(\hat{\text{Var}}(\log(\hat{\gamma})))^{1/2}), \text{ and } \gamma_{\text{u}} = \delta(\underline{\text{n}}_E, \text{n}^{(\text{C})}_{1+})(\hat{\gamma} + \text{Z}_{\text{a}/2}(\hat{\text{Var}}(\hat{\gamma}))^{1/2})$$
$$+ (1-\delta(\underline{\text{n}}_E, \text{n}^{(\text{C})}_{1+}))\hat{\gamma}\exp(\text{Z}_{\text{a}/2}(\hat{\text{Var}}(\log(\hat{\gamma})))^{1/2}).$$

On the basis of some empirical evaluations of the coverage probability and average length of Eq. (10) based on simulations as stated in the following section, we recommend the constant K to be set equal to 2.5 when using Eq. (10).

3. Monte Carlo simulation

To evaluate the finite sample performance of interval estimators Eqs. (3), (4), (7), (9), and (10), we employ Monte Carlo simulation. Given the probability that a randomly selected patient who would comply with the experimental treatment $\theta(=p_{+1}^{(E)}=p_{+1}^{(C)})$ and the conditional probability of response $p_{1|1}^{(C)}(=p_{11}^{(C)}/p_{+1}^{(C)})$ given the patient who would comply with the experimental treatment (if he/she were actually assigned to receive this therapy) in the control group, we can uniquely determine the cell probability $p_{11}^{(C)}=p_{1|1}^{(C)}\theta$. For a given value of γ, we can further determine the cell probability: $p_{11}^{(E)}=\gamma p_{11}^{(C)}$. Similarly, when the conditional probability $p_{1|0}^{(C)}$ of response among patients who would not comply with receiving the experimental treatment in the control group is given, we can determine the cell probability $p_{10}^{(C)}(=p_{10}^{(E)})$ by $p_{1|0}^{(C)}(1-\theta)$. We arbitrarily choose $p_{1|0}^{(C)}=4p_{1|1}^{(C)}/3$. For simplicity, we concentrate our attention on equal sample allocation (i.e., $n_E=n_C$). We consider the situations, in which the probability of a randomly selected patient who would comply with the experimental treatment $\theta(=p_{+1}^{(C)}=p_{+1}^{(C)})=0.50,\ 0.80$; the underlying RR of proportions of response between two comparison groups among patients who would comply with the experimental treatment $\gamma=1.0,\ 1/2,\ 1/3$; the conditional probability of responses, given the patient who would comply with the experimental treatment in the control group $p_{1|1}^{(C)}=0.30,\ 0.50$; and the number of patients assigned to either of treatment groups $n\ (=n_E=n_C)=30,\ 50,\ 100$. For each configuration determined by a combination of these parameters, we apply SAS [16] and generate 10,000 repeated samples of observations following the desired multinomial distributions to calculate the estimated coverage probability and average length for each interval estimator. Note that if a simulated sample led to $\hat{\gamma}=0$ or ∞, we would not be able to apply any interval estimator considered here. Furthermore, if $A<0$ or $B^2-AC<0$, as noted previously, we could not apply interval estimator Eq. (7) either. When calculating the coverage probability of each interval estimator, we are conditional upon the simulated samples where a 95% confidence interval exists. The coverage probability is then calculated as the proportion of these samples that the resulting 95% confidence interval includes the underlying true value of γ. Similarly, the average length is calculated as the average of the length of the resulting 95% confidence interval over those samples for which the interval estimate exists. For completeness, we also calculate the probability of failure to apply an interval estimator as the proportion out of 10,000 simulated samples for which the 95% confidence interval using this interval estimator does not exist.

4. Results

Table 1 summarizes the estimated coverage probability and average length (in parenthesis) of the 95% confidence interval using interval estimators Eqs. (3), (4), (7), (9), and (10). To help readers easily see whether a given interval estimator is accurate or not, we mark by "×" the estimated coverage probabilities, when they are less than the desired 95% confidence level by more than 1% in Table 1. First, we can see that the coverage probability of 95% confidence intervals using Eqs. (3) and (9) tends to be less than the desired 95% confidence level. For example, when $\theta=0.50$, $\gamma=0.50$, $p_{1|1}^{(C)}=0.30$, and $n=50$, the coverage probabilities of 95% confidence intervals using Eqs. (3) and (9) are 0.888 and 0.854, respectively. Using Eq. (7) may generally improve the estimated coverage probability compared to Eqs. (3) and (9). We also note that the coverage probability of using interval estimators Eqs. (4) and (10) is consistently larger than or equal to the 95% confidence level in all the situations in Table 1. However, using the former can substantially lose efficiency as compared with using the latter. For example, when $\theta=0.50$, $\gamma=1.00$, $p_{1|1}^{(C)}=0.30$, and $n=30$, the estimated average length of 95% confidence interval using Eq. (4) is 241.8, that is approximately 40 times of that using Eq. (10). From Table 1, we also find that using the interval estimator Eq. (10) consistently produces the most efficient 95% confidence interval among interval estimators subject to the estimated coverage probability ≥ the desired 95% confidence level. Table 2 shows that when there is a low proportion of patients who would comply with the experimental treatment (i.e., $\theta=0.50$) and the number of patients n is not large (≤ 50), the probability of failure to apply interval estimators Eqs. (3), (4), (9), and (10) due to the non-existence of the MLE $\hat{\gamma}$ or the probability of failure to apply interval estimator Eq. (7) due to this and the additionally required conditions for finding the two distinct real roots of the quadratic Eq. (6) can be non-negligible. When θ is large ($\theta\geq 0.80$) or when the number of patients is

Table 1
The estimated coverage probability and average length (in parenthesis) of 95% confidence interval using interval estimators Eqs. (3), (4), (7), (9), and (10) in the situations where the probability of a randomly selected patient who would comply with the experimental treatment $\theta = 0.50, 0.80$; the underlying RR of probabilities of response between two comparison groups among patients who would comply with the experimental treatment $\gamma = 1, 1/2, 1/3$; the conditional probability of response, given the patient who would comply with the experimental treatment in the control group $p_{1|1}^{(C)} = 0.30, 0.50$; and the number of patients per group $n\ (= n_T) = 30, 50, 100$

θ	γ	$p_{1\|1}^{(C)}$	n	Eq. (3)	Eq. (4)	Eq. (7)	Eq. (9)	Eq. (10)
0.50	1.00	0.30	30	0.855×	0.971	0.908×	0.811×	0.955√
				(5.213)	(241.8)	(5.997)	(3.586)	(6.092)
			50	0.884×	0.963	0.935×	0.851×	0.963√
				(5.595)	(1256)	(10.30)	(3.971)	(6.517)
			100	0.907×	0.962	0.964	0.887×	0.962√
				(4.071)	(13069)	(6.442)	(3.198)	(4.842)
		0.50	30	0.905×	0.968	0.946	0.839×	0.967√
				(4.003)	(235.0)	(3.451)	(2.643)	(4.758)
			50	0.919×	0.966	0.962	0.870×	0.966√
				(3.062)	(830.3)	(4.230)	(2.262)	(3.624)
			100	0.932×	0.964	0.962	0.901×	0.964√
				(1.416)	(3.092)	(3.043)	(1.319)	(1.594)
	0.50	0.30	30	0.899×	1.000	0.928×	0.868×	0.957√
				(2.983)	(142.8)	(3.340)	(2.061)	(3.527)
			50	0.888×	0.991	0.924×	0.854×	0.959√
				(2.944)	(692.0)	(5.712)	(2.096)	(3.514)
			100	0.909×	0.977	0.951	0.891×	0.976√
				(2.072)	(7199)	(3.391)	(1.646)	(2.527)
		0.50	30	0.900×	0.994	0.932×	0.856×	0.979√
				(2.094)	(110.0)	(1.976)	(1.412)	(2.658)
			50	0.916×	0.983	0.951	0.871×	0.980√
				(1.632)	(412.1)	(2.374)	(1.222)	(2.035)
			100	0.930×	0.972	0.954	0.900×	0.972√
				(0.882)	(3322)	(1.658)	(0.794)	(1.001)
	0.33	0.30	30	0.971	1.000	0.975	0.947	1.000√
				(2.293)	(111.5)	(2.722)	(1.594)	(2.699)
			50	0.909×	1.000	0.937×	0.882×	0.967√
				(2.151)	(507.6)	(3.609)	(1.534)	(2.591)
			100	0.903×	0.992	0.945	0.886×	0.976√
				(1.397)	(4843)	(2.499)	(1.122)	(1.766)
		0.50	30	0.935×	0.995	0.946	0.882×	0.995√
				(1.546)	(80.67)	(1.485)	(1.047)	(1.999)
			50	0.915×	0.992	0.941	0.885×	0.984√
				(1.115)	(219.6)	(1.675)	(0.854)	(1.467)
			100	0.922×	0.973	0.948	0.904×	0.973√
				(0.657)	(2187)	(1.217)	(0.591)	(0.768)
0.80	1.00	0.30	30	0.906×	0.977	0.958	0.893×	0.975√
				(3.167)	(32.92)	(6.409)	(2.707)	(4.045)
			50	0.927×	0.974	0.962	0.917×	0.974√
				(2.169)	(31.62)	(6.760)	(2.001)	(2.617)
			100	0.943	0.964	0.949	0.938×	0.964√
				(1.241)	(1.542)	(1.825)	(1.219)	(1.333)
		0.50	30	0.932×	0.972	0.960	0.910×	0.972√
				(1.581)	(2.941)	(3.097)	(1.471)	(1.808)
			50	0.939×	0.963	0.947	0.923×	0.963√
				(1.105)	(1.181)	(1.602)	(1.074)	(1.170)
			100	0.946	0.954√	0.943	0.940	0.954√
				(0.730)	(0.746)	(0.801)	(0.722)	(0.746)
	0.50	0.30	30	0.908×	0.995	0.945	0.899×	0.978√
				(1.805)	(17.17)	(3.569)	(1.552)	(2.443)
			50	0.921×	0.984	0.945	0.904×	0.982√
				(1.228)	(9.513)	(3.389)	(1.143)	(1.571)
			100	0.931×	0.962	0.945	0.924×	0.962√

(continued on next page)

Table 1 (*continued*)

θ	γ	$p_{1\|1}^{(C)}$	n	Eq. (3)	Eq. (4)	Eq. (7)	Eq. (9)	Eq. (10)
				(0.743)	(0.823)	(1.044)	(0.731)	(0.818)
		0.50	30	0.922×	0.977	0.947	0.914×	0.976√
				(1.003)	(3.233)	(1.766)	(0.925)	(1.210)
			50	0.940	0.962	0.942	0.929×	0.962√
				(0.724)	(0.799)	(1.074)	(0.703)	(0.792)
			100	0.946	0.958√	0.947	0.938×	0.958√
				(0.483)	(0.502)	(0.522)	(0.477)	(0.502)
	0.33	0.30	30	0.947	0.991	0.968	0.923×	0.991√
				(1.361)	(15.56)	(2.446)	(1.166)	(1.885)
			50	0.911×	0.987	0.943	0.905×	0.980√
				(0.911)	(13.73)	(2.522)	(0.839)	(1.226)
			100	0.924×	0.965	0.943	0.922×	0.965√
				(0.567)	(0.646)	(0.795)	(0.557)	(0.643)
		0.50	30	0.920×	0.977	0.933×	0.899×	0.977√
				(0.752)	(1.756)	(1.270)	(0.693)	(0.979)
			50	0.921×	0.966	0.936×	0.915×	0.966√
				(0.560)	(0.671)	(0.745)	(0.541)	(0.639)
			100	0.938×	0.956√	0.942	0.934×	0.956√
				(0.383)	(0.405)	(0.411)	(0.378)	(0.405)

× means the estimated coverage probability is less than the desired 95% confidence level by >1%.
√ indicates that the interval estimate has the shortest estimated average length among 95% confidence intervals with the estimated coverage probability larger than or equal to the desired 95% confidence level.

large (=100), except for using Eq. (7), the probability of failure to produce a 95% confidence interval using Eqs. (3), (4), (9), and (10) is generally small or negligible.

5. An example

To illustrate the use of interval estimators Eqs. (3), (4), (7), (9), and (10), consider the study of taking vitamin A supplementation to reduce the mortality rate among preschool children in rural Indonesia [1]. Children were randomly assigned to either the treatment group of receiving a large oral dose of vitamin A for 2 to 3 months following baseline enumeration and again 6 months later, or to the control group of receiving no vitamin A supplementation. In the control group, because children were precluded from receiving a placebo for ethic reasons, we could observe only the total number of survival children without the information on children who would comply with vitamin A if they were actually assigned to receive this vitamin supplementation. As shown elsewhere [2], we had the number of deaths in the control group, $(n_{1+}^{(C)}=)74$ out of $(n_C=)11{,}588$ children. Furthermore, we had $n_{11}^{(E)}=12$, $n_{10}^{(E)}=34$, $n_{01}^{(E)}=9663$, $n_{00}^{(E)}=2385$ for the total number $(n_E=)12{,}094$ of children assigned to the treatment of receiving vitamin A supplementation. Suppose that we are interested in estimating the RR of mortality rates between the experimental group receiving vitamin A supplementation and the control group receiving no vitamin A supplementation among children who would comply with taking vitamin A supplementation. Given the above data, we obtain the MLE $\hat{\gamma}=0.278$. When applying interval estimators Eqs. (3), (4), (7), (9), and (10), we obtain the 95% confidence intervals as [0.071, 0.484], [0.132, 0.584], [0.112, 0.613], [0.071, 0.484], and [0.132, 0.584], respectively. We can see that the resulting interval estimate using Eq. (4) tends to shift to the right as compared with Eqs. (3) and (9). We can also see that, as what we would expect, interval estimators Eqs. (3) and (9) (of which both are derived directly from large sample properties of the MLE) are the same with such a large sample size as given here. Furthermore, because the length of Eq. (4) is not greater than 2.5 times of the length of Eq. (3), the interval estimate Eq. (10) is, by definition, identical to that of Eq. (4). Finally, we note that since all upper limits fall below 1, we may conclude that there is a significant evidence to support that the usage of vitamin A supplementation can reduce the mortality rate among preschool children at 5% level.

To investigate whether it is appropriate to use Eqs. (3), (4), (7), (9), and (10) in the particular situations considered in this example, we may apply Monte Carlo simulation with parameter values determined by the empirical estimates from the data. We have generated 10,000 repeated samples of observations $\underline{n}_E$ following the multinomial distribution with parameters $n_E=12{,}094$ and $p_{11}^{(E)}=0.0010$, $p_{10}^{(E)}=0.0028$, $p_{01}^{(E)}=0.7990$, $p_{00}^{(E)}=0.1972$ and

Table 2
The estimated probability of failure to apply interval estimators Eqs. (3), (4), (7), (9), and (10) to produce 95% confidence interval in the situations where the probability of a randomly selected patient who would comply with the experimental treatment $\theta=0.50, 0.80$; the underlying RR of probabilities of response between two comparison groups among patients who would comply with the experimental treatment $\gamma=1, 1/2, 1/3$; the conditional probability of response given the patient who would comply with the experimental treatment in the control group $p_{1|1}^{(C)}=0.30, 0.50$; and the number of patients per group n $(=n_T)=30, 50, 100$

θ	γ	$p_{1\|1}^{(C)}$	n	Eq. (3)†	Eq. (4)†	Eq. (7)	Eq. (9)†	Eq. (10)†
0.50	1.00	0.30	30	0.126	0.126	0.719	0.126	0.126
			50	0.056	0.056	0.594	0.056	0.056
			100	0.013	0.013	0.333	0.013	0.013
		0.50	30	0.032	0.032	0.491	0.032	0.032
			50	0.007	0.007	0.265	0.007	0.007
			100	0.000	0.000	0.047	0.000	0.000
	0.50	0.30	30	0.200	0.200	0.745	0.200	0.200
			50	0.076	0.076	0.588	0.076	0.076
			100	0.011	0.011	0.326	0.011	0.011
		0.50	30	0.051	0.051	0.507	0.051	0.051
			50	0.011	0.011	0.265	0.011	0.011
			100	0.000	0.000	0.045	0.000	0.000
	0.33	0.30	30	0.301	0.301	0.771	0.301	0.301
			50	0.127	0.127	0.618	0.127	0.127
			100	0.015	0.015	0.330	0.015	0.015
		0.50	30	0.099	0.099	0.527	0.099	0.099
			50	0.019	0.019	0.274	0.019	0.019
			100	0.000	0.000	0.047	0.000	0.000
0.80	1.00	0.30	30	0.011	0.011	0.296	0.011	0.011
			50	0.001	0.001	0.112	0.001	0.001
			100	0.000	0.000	0.006	0.000	0.000
		0.50	30	0.000	0.000	0.060	0.000	0.000
			50	0.000	0.000	0.006	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000
	0.50	0.30	30	0.029	0.029	0.315	0.029	0.029
			50	0.002	0.002	0.114	0.002	0.002
			100	0.000	0.000	0.006	0.000	0.000
		0.50	30	0.002	0.002	0.063	0.002	0.002
			50	0.000	0.000	0.005	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000
	0.33	0.30	30	0.091	0.091	0.347	0.091	0.091
			50	0.016	0.016	0.121	0.016	0.016
			100	0.000	0.000	0.005	0.000	0.000
		0.50	30	0.015	0.015	0.074	0.015	0.015
			50	0.000	0.000	0.006	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000

† The probability of failure to apply Eqs. (3), (4), (9) and (10) to produce 95% confidence interval is actually equal to the probability of obtaining the MLE $\hat{\gamma}=0$ or ∞.

observations $\underline{n}_C=(n_{11}^{(C)},n_{10}^{(C)},n_{01}^{(E)},n_{00}^{(C)})'$ (following the multinomial distribution with parameters $n_C=11{,}588$ and $p_{11}^{(C)}=0.0036$, $p_{10}^{(C)}=0.0028$, $p_{01}^{(C)}=0.7964$, $p_{00}^{(C)}=0.1972$. Note that in the placebo group, we can only observe the marginal total $n_{1+}^{(C)}$, while the random vector of observation $\underline{n}_C$ is not observable. Based on the simulation, we have found that the probability of failure to produce a 95% confidence interval to be 0.013 when using Eq. (7) and 0.000 when using all the other interval estimators. When applying Eqs. (3), (4), (7), (9), and (10), we have obtained the estimated coverage probability and average length (in parenthesis) of the 95% confidence intervals to be 0.925 (0.475), 0.967 (0.561), 0.950 (0.822), 0.925 (0.475) and 0.967 (0.538), respectively. These suggest that the coverage probability of using Eqs. (3) and (9) tends to be less than the desired 95% confidence level, while interval estimators Eqs. (4), (7) and (10) are all appropriate for use based on the estimated coverage probability. We also find that the interval estimator Eq. (10) is the most efficient among the latter three interval estimators. This is certainly consistent with our previous findings in Table 1.

6. Discussion

Because it has been successfully employed to produce an interval estimate for the RR in many other situations [14,17–22], using the logarithmic transformation of the MLE is probably the most commonly-used method when we want to produce a confidence interval for a ratio of two parameters. In fact, the interval estimator Eq. (4) for the RR using the logarithmic transformation of the MLE was also proposed elsewhere [2] for the SCRT. However, when the number of patients in both comparison groups is not large, the sampling distribution of the MLE $\hat{\gamma}$ Eq. (1) can be skewed to the right and hence the resulting interval estimate Eq. (4) may have, as noted previously, a quite wide confidence interval in this case. For example, consider the situation in which $\theta=0.50$, $\gamma=0.30$, $p_{1|1}^{(C)}=0.30$ and $n=30$ in Table 1. In our simulated samples, we obtained the following sample data: $\underline{n}_E=(5, 6, 14, 5)$ and $n_{1+}^{(C)}=7$. These give the MLE $\hat{\gamma}=5$ with an estimated variance $\hat{\mathrm{Var}}(\hat{\gamma})=248.33$. Furthermore, the 95% confidence intervals using Eqs. (3), (4), (9), and (10) are: [0.000, 35.887], [0.010, 2408.615], [0.000, 21.599] and [0.000, 35.887], while the interval estimate using Eq. (7) does not exist based on these data. Note that the above resulting interval estimate Eq. (4) is approximately 67 times of the length of Eq. (3). This explains the reason why the average length of Eq. (4) using the logarithmic transformation can be, as found in Table 1, much larger than those of the others when n is small or moderate. We also note that except for using Eq. (4), the estimated average length of Eqs. (3), (9), (7), and (10) generally decreases as the number of patients in both comparison groups increases. The irregular pattern of the estimated average length of using Eq. (4) is actually due to the fact that as n increases in the situations where both $p_{11}^{(T)}$ and $p_{1+}^{(C)}-p_{10}^{(T)}$ are small, we increase the chance of obtaining samples in which $\hat{\mathrm{p}}_{11}^{(T)}$ and $\hat{\mathrm{p}}_{1+}^{(C)}-\hat{\mathrm{p}}_{10}^{(T)}$ possess small positive values instead of samples in which $\hat{\mathrm{p}}_{11}^{(T)}=0$ or $\hat{\mathrm{p}}_{1+}^{(C)}-\hat{\mathrm{p}}_{10}^{(T)}\leq 0$. These later samples are excluded from the denominator due to the non-existence of the MLE when calculating the estimated average length. Thus, the probability of obtaining a large estimated variance $\hat{\mathrm{Var}}(\log(\hat{\gamma}))$ (instead of obtaining an undefined estimated variance) can increase as n increases in these situations. This accounts for the reason why the estimated average length of Eq. (4) may increase as n increases when the probability of failure to apply this interval is non-negligible. However, when the number of patients reaches a large value (say, $n\geq 400$) so that the probability of failure to produce a confidence interval using Eq. (4) is negligible, the estimated average length of Eq. (4) will decrease as n increases as well.

Note that an ideal interval estimator is the one which can produce a $100(1-\alpha)$ percent confidence interval with the coverage probability consistently $\geq$ the desired confidence level, and with the average length uniformly shorter than any other interval estimator for which the coverage probability is also $\geq$ the desired confidence level. From Table 1, we can see that the interval estimator Eq. (10) can consistently perform well even when the number of patients in both treatments is as small as $n=30$ and uniformly outperform all the other estimators in almost all the situations considered here. Note also that when $\theta=0.50$, $\gamma=1/3$, $p_{1|1}^{(C)}=0.30$, and $n=30$, the expected number of patients in the cell of $p_{11}^{(E)}$ is less than 2 patients. Thus, it is really of no surprise to see that the probability of failure to apply any interval estimator due to the non-existence for the MLE $\hat{\gamma}$ can be large in this extreme case (Table 2). In practice, however, we rarely draw an inference based on only a single or two compliance patients with positive response in the experimental treatment group. Therefore, the large probability of failure to apply interval estimators in this case should be interpreted as an inherent limitation in data rather than an issue of concern in application of interval estimators discussed here. On the other hand, when $\theta=0.80$ and $n=50$, the probability of failure to apply interval estimators Eqs. (3), (4), (9), and (10) are generally small or even negligible (Table 2). Therefore, interval estimator Eq. (10) should be useful and applicable under most situations which we will encounter in practice.

Note that whenever both $\hat{p}_{11}^{(E)}>0$ and $\hat{p}_{1+}^{(C)}-\hat{p}_{10}^{(E)}>0$, the MLE $\hat{\gamma}$ exists and all interval estimators Eqs. (3), (4), (9), and (10) are applicable. In other words, the coverage probability and the average length for Eqs. (3), (4), (9), and (10) are calculated over the same set of sample spaces. Therefore, the comparison of performance between these interval estimators would not be confounded by excluding samples for which $\hat{p}_{11}^{(E)}=0$ or $\hat{p}_{1+}^{(C)}-\hat{p}_{10}^{(E)}\leq 0$. On the other hand, when comparing (7) with the others, we may need to account for the results in Table 2. Note also that when the confidence interval using (7) exists, interval estimator (7) can outperform interval estimators (3) and (9) (Table 1). This is because interval estimator Eq. (7) is derived from $Z(\gamma)=\hat{p}_{11}^{(E)}-\gamma(\hat{p}_{1+}^{(C)}-\hat{p}_{10}^{(E)})$, of which the sampling distribution is likely to be less skewed than that of the MLE $\hat{\gamma}$, on which interval estimators Eqs. (3) and (9) are based. On the other hand, to employ interval estimator Eq. (7), our data are required to meet both conditions $A>0$ and $B^2-AC>0$ for solving the quadratic Eq. (6). In fact, we can easily see that the condition $A>0$ is equivalent to $(\hat{p}_{1+}^{(C)}-\hat{p}_{10}^{(E)})^2/[\hat{p}_{1+}^{(C)}(1-\hat{p}_{1+}^{(C)})/n_C+\hat{p}_{10}^{(E)}(1-\hat{p}_{10}^{(C)})/n_E]\geq Z_{\alpha/2}^2$. In other words, interval estimator Eq. (7) would be inapplicable if we could not reject the null hypothesis $H_0: p_{1+}^{(C)}-p_{10}^{(C)}$ (which is the denominator of γ) $=0$ at α-level. When the number of patients n_T

is small or when the underlying probability $p_{1+}^{(C)} - p_{10}^{(C)}$ is close to 0, we all know that the power of rejecting such a null hypothesis is likely to be small and so is the probability of obtaining $A > 0$. This can partially account for the reason why the probability of failure to apply interval Eq. (7) is high, especially when the number of patients $n(= n_T)$ is not large (Table 2).

The interval estimator Eq. (10) can, as noted previously, perform well even when the number of patients is 30 per treatment in a variety of situations (Table 1). When the number of patients per treatment is large, we expect that the probability of the estimated average length of Eq. (4) larger than that of Eq. (3) by >2.5 times is close to 0, and hence interval estimator Eq. (10) will essentially have the same asymptotic behavior as the most frequently-used interval estimator Eq. (4) using the logarithmic transformation.

In summary, we consider and develop five interval estimators for the RR under the SCRT. We apply Monte Carlo simulation to evaluate the performance of these interval estimators. We note that the interval estimator Eq. (3) using Wald's statistic and the interval estimator Eq. (9) directly derived from the MLE can lose accuracy. We find that the commonly-used interval estimator using the logarithmic transformation Eq. (4) proposed elsewhere can lose efficiency. We also find that the probability of failure to apply the interval estimator Eq. (7) derived from an idea used in Fieller's Theorem can be non-negligible when the number of patients per treatment is not large. Finally, we find that the interval estimator Eq. (10) can consistently perform well even when the number of patients is moderate, and hence we recommend this estimator for general use. The results, the findings, and the discussions should have use for biostatisticians or statistically literate clinicians when they encounter a SCRT.

Acknowledgements

The author wishes to thank the two reviewers for many valuable comments to improve the contents of this paper. The author also wants to thank Dr. D. Sandweiss for his editorial suggestions and help to improve the clarity of this paper.

References

[1] Sommer A, Tarwotjo I, Djunaedi E, et al. Impact of vitamin A supplementation on childhood mortality: a randomized controlled community trial. Lancet 1986;i:1169–73.
[2] Sommer A, Zeger SL. On estimating efficacy from clinical trials. Stat Med 1991;10:45–52.
[3] Casella G, Berger RL. Statistical Inference. Duxbury Belmont, CA: Duxbury; 1990.
[4] Zelen M. A new design for randomized clinical trials. N Engl J Med 1979;300:1242–5.
[5] Ellenberg SS. Randomization designs in comparative clinical trials. N Engl J Med 1984;310:1404–8.
[6] Bernhard G, Compagnone D. Binary data in prerandomized designs. Biom J 1989;31:29–33.
[7] Anbar D. The relative efficiency of Zelen's prerandomization design for clinical trials. Biometrics 1983;39:711–8.
[8] McHugh R. Validity and treatment dilution in Zelen's single consent design. Stat Med 1984;3:215–8.
[9] Matts J, McHugh R. Randomization and efficiency in Zelen's single consent design. Biometrics 1987;43:885–94.
[10] Matts JP, McHugh RB. Precision estimation in Zelen's single-consent design. Biom J 1993;35:65–72.
[11] Brunner E, Neumann N. On the mathematical basis of Zelen's prerandomized designs. Methods Inf Med 1985;24:120–30.
[12] Zelen M. Randomized consent designs for clinical trials: an update. Stat Med 1990;9:645–56.
[13] Zelen M. Response. J Chronic Dis 1986;39:247–9.
[14] Katz D, Baptista J, Azen SP, Pike MC. Obtaining confidence intervals for the risk ratio in cohort studies. Biometrics 1978;34:469–74.
[15] Lui K-J. Estimation of attributable risk for case-control studies with multiple matching. Stat Med 2005;24:2953–62.
[16] SAS Institute Inc. SAS Language, Reference Version 6. 1st ed. Cary, North Carolina: SAS Institute; 1990.
[17] Fleiss JL. Statistical Methods for Rates and Proportions. 2nd ed. New York: Wiley; 1981.
[18] Lui K-J. Statistical Estimation of Epidemiological Risk. New York: Wiley; 2004.
[19] Greenland S, Robins JM. Estimation of a common effect parameter from sparse follow-up data. Biometrics 1985;41:55–68.
[20] Lui K-J. Interval estimation of proportion ratio under multiple matching. Stat Med 2005;24:1275–85.
[21] Lui K-J. Eight interval estimators of a common rate ratio under stratified Poisson sampling. Stat Med 2004;23:1283–96.
[22] Lui K-J. A Monte Carlo evaluation of five interval estimators for the relative risk in sparse data. Biom J 2006;48:131–43.