International Statistical Review

International Statistical Review (2023) doi: 10.1111/insr.12539

Estimating the Reciprocal of a Binomial Proportion

Jiajin Wei[1], Ping He[2] and Tiejun Tong[1]

[1]*Department of Mathematics, Hong Kong Baptist University, Hong Kong*
[2]*Faculty of Science and Technology, BNU-HKBU United International College, Zhuhai, China*
Corresponding to: Ping He, Faculty of Science and Technology, BNU-HKBU United International College, Zhuhai, China Email: heping@uic.edu.cn
Tiejun Tong, Department of Mathematics, Hong Kong Baptist University, Hong Kong. Email: tongt@hkbu.edu.hk

Summary

The binomial proportion is a classic parameter with many applications and has also been extensively studied in the literature. By contrast, the reciprocal of the binomial proportion, or the inverse proportion, is often overlooked, even though it also plays an important role in various fields. To estimate the inverse proportion, the maximum likelihood method fails to yield a valid estimate when there is no successful event in the Bernoulli trials. To overcome this zero-event problem, several methods have been introduced in the previous literature. Yet to the best of our knowledge, there is little work on a theoretical comparison of the existing estimators. In this paper, we first review some commonly used estimators for the inverse proportion, study their asymptotic properties, and then develop a new estimator that aims to eliminate the estimation bias. We further conduct Monte Carlo simulations to compare the finite sample performance of the existing and new estimators, and also apply them to handle the zero-event problem in a meta-analysis of COVID-19 data for assessing the relative risks of physical distancing on the infection of coronavirus.

Key words: binomial proportion; inverse proportion; relative risk; shrinkage estimator; zero-event problem.

1 Introduction

The binomial distribution is one of the most important distributions in statistics, which has been extensively studied in the literature with a wide range of applications. This classical distribution has two parameters n and p, where n is the number of independent Bernoulli trials and p is the probability of success in each trial (Hogg *et al.*, 2005). The probability of success, p, is also referred to as the binomial proportion. For excellent reviews on its estimation and inference, one may refer to, for example, Agresti & Coull (1998) and Brown *et al.* (2001).

Apart from the parameter p, it is known that some of its functions, say $p(1-p)$ and $\ln(p)$, also play important roles in statistics and have received much attention. In this article, we are interested in the reciprocal function

$$\theta = \frac{1}{p}, \tag{1}$$

which is another important function of p yet is often overlooked in the literature. For convenience, we also refer to θ in formula (1) as the inverse proportion of the binomial distribution. To demonstrate its usefulness, we will introduce some motivating examples in Section 2 that connect the inverse proportion with the relative risk (RR) and with the Horvitz–Thompson estimator (Horvitz & Thompson., 1952; Fattorini, 2006). Moreover, we will also introduce in Section 6 a relationship of the inverse proportion to the number needed to treat (NNT) and the reduction in number to treat (RNT) in clinical studies, and present some future directions (Laupacis *et al.*, 1988; Altman, 1998; Hutton, 2000; Zhang & Yin, 2021).

To start with, let $X = \sum_{i=1}^{n} X_i$, where X_i are independent and identically distributed random variables from a Bernoulli distribution with success probability $p \in (0, 1)$. Then equivalently, X follows a binomial distribution with parameters $n \geq 1$ and p. Now if we want to estimate the inverse proportion θ, a simple method will be to apply the maximum likelihood estimation (MLE) and it yields

$$\hat{\theta}_{\mathrm{MLE}} = \frac{n}{X}. \tag{2}$$

This estimator is, however, not a valid estimator because it is not defined when $X = 0$, that is, when there is no successful event in n trials. We refer to this problem as *the zero-event problem* in the point estimation of θ. In fact, the same problem also exists in the interval estimation of p. Specifically by Hogg *et al.* (2005), the $100(1 - \alpha)\%$ Wald interval is given as

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}},$$

where $\hat{p} = X/n$, and $z_{\alpha/2}$ is the upper $\alpha/2$ percentile of the standard normal distribution. When $X = 0$, the lower and upper limits of the Wald interval are both zero; and consequently, they will not be able to provide a $100(1 - \alpha)\%$ coverage probability for the true proportion.

To overcome the zero-event problem, Hanley & Lippman-Hand (1983) proposed the 'Rule of Three' to approximate the upper limit of the 95% confidence interval (CI) for p. Specifically, as the upper limit of the one-sided CI for p is $1 - 0.05^{1/n}$ when $X = 0$, the authors approximated this upper limit by $3/n$, which then yields the simplified CI as $(0, 3/n)$. For more discussion on the 'Rule of Three', one may refer to Tuyl *et al.* (2009) and the references therein. In particular, we note that the Wilson interval (Wilson, 1927) and the Agresti–Coull interval (Agresti & Coull, 1998) for p have also been referred to as the variations of the 'Rule of Three'.

The Wilson interval was originated from Laplace who proposed the 'Law of Succession' in the 18th century. As mentioned in Good (1980), Laplace's estimator for the binomial proportion was given as $(X + 1)/(n + 2)$, which is indeed a shrinkage estimator for p. Wilson (1927) generalised the shrinkage idea and proposed an updated 'Law of Succession' as $\tilde{p}(c) = (X + c)/(n + 2c)$, where $c > 0$ is a regularisation parameter. By applying the Wilson estimator, Agresti & Coull (1998) substituted $\tilde{p}(c)$ for $\hat{p}$ in the Wald interval, which yields the Agresti–Coull interval

$$\tilde{p}(c) \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(c)[1 - \tilde{p}(c)]}{n}}.$$

It is also noteworthy that the Agresti–Coull interval always performs better than the Wald interval, no matter whether n is large or small (Brown *et al.*, 2001).

By applying the Wilson estimator $\tilde{p}(c)$, one may estimate the inverse proportion as

$$\tilde{\theta}(c) = \frac{n+2c}{X+c}, \quad c > 0. \tag{3}$$

Note that the estimator with form (3) does not suffer from the zero-event problem, and so it provides a valid estimate of θ for any given $c > 0$. In particular, two special cases of estimator (3) with $c = 0.5$ and 1 have been widely applied in the previous literature (Walter, 1975; Carter *et al.*, 2010). Moreover, there are other estimators that follow the structure of (3) including, for example, a piecewise estimator (PE) with corrections only on $X = 0$ or n (Schwarzer, 2007). In addition to (3), another family of shrinkage estimators for the inverse proportion takes the form of

$$\hat{\theta}(c) = \frac{n+c}{X+c}, \quad c > 0. \tag{4}$$

For the special case $\hat{\theta}(0.5)$, it has been investigated by Pettigrew *et al.* (1986) and Hartung & Knapp (2001). More recently, Fattorini (2006) applied $\hat{\theta}(1)$ to estimate θ in sampling designs and demonstrated that it provides an asymptotically unbiased estimator of θ as n tends to infinity. For more results on $\hat{\theta}(c)$, see also Chao & Strawderman (1972), Gamrot (2013), Seber (2013), and the references therein.

The remainder of this paper is organised as follows. In Section 2, we briefly review the literature and introduce two motivating examples where an estimate of the inverse proportion is desired. In Section 3, we first compare the theoretical properties of the existing estimators, and then derive the optimal estimator in family (4) that minimises the estimation bias. In Section 4, we conduct Monte Carlo simulations to compare the existing and new estimators in terms of the relative bias, Stein loss and mean squared error. In Section 5, we further apply our new method to handle the zero-event problem in a meta-analysis of COVID-19 data for assessing the relative risks of physical distancing on the infection of coronavirus. Lastly, we conclude the paper in Section 6 with some discussion and future work, and present the Appendices in the supporting information.

2 MOTIVATING EXAMPLES

In this section, we provide two motivating examples in which an accurate estimate of the inverse proportion θ is highly desired.

2.1 Relative Risk

In clinical studies, the relative risk (RR), also known as the risk ratio, is a commonly used effect size for measuring the effectiveness of a treatment or intervention (Agresti, 2003; Wei *et al.*, 2021). Specifically, RR is defined as

$$\mathrm{RR} = \frac{p_1}{p_2}, \tag{5}$$

where p_1 is the event probability in the exposed group, and p_2 is the event probability in the unexposed group.

To estimate RR, we assume that there are n_1 samples in the exposed group with X_1 being the number of events, and n_2 samples in the unexposed group with X_2 being the number of events. Let also X_1 follow a binomial distribution with parameters n_1 and p_1, X_2 follow a binomial distribution with parameters n_2 and p_2, and that they are independent of each other. Then by (5) and applying the MLEs of p_1 and p_2 respectively, RR can be estimated by

$$\widehat{\mathrm{RR}} = \frac{X_1/n_1}{X_2/n_2} = \frac{X_1 n_2}{X_2 n_1}. \tag{6}$$

A problem of this estimator is, however, that it suffers from the zero-event problem when $X_2 = 0$. To overcome this problem, there are a few popular suggestions in the literature to further improve the RR estimator in (6).

(i) Walter (1975) introduced a modified estimator of RR as $\tilde{\mathrm{RR}}(0.5) = (X_1 + 0.5)(n_2 + 1)/[(X_2 + 0.5)(n_1 + 1)]$. Following this idea, the inverse proportion of the unexposed group is, in fact, estimated by the Walter estimator

$$\tilde{\theta}(0.5) = \frac{n_2 + 1}{X_2 + 0.5}, \tag{7}$$

which is a special case of estimator (3) with $c = 0.5$.

(ii) Pettigrew *et al.* (1986) proposed to estimate p_i by $(X_i + 0.5)/(n_i + 0.5)$ for $i = 1$ or 2, and further concluded that $\ln[(X_i + 0.5)/(n_i + 0.5)]$ is an unbiased estimator of $\ln(p_i)$ by ignoring the term $O(n^{-2})$. Accordingly, the Pettigrew estimator for the inverse proportion can be given as

$$\hat{\theta}(0.5) = \frac{n_2 + 0.5}{X_2 + 0.5}, \tag{8}$$

which is a special case of estimator (4) with $c = 0.5$.

(iii) Originated from (3), a family of piecewise estimators is defined as

$$\tilde{\theta}_{\mathrm{PE}}(c) = \frac{n + 2cI(X = 0\,\mathrm{or}\,n)}{X + cI(X = 0\,\mathrm{or}\,n)}, \quad c > 0, \tag{9}$$

where $I(\cdot)$ is the indicator function. Particularly, one special case with $c = 0.5$ that has been extensively applied in clinical studies (Carter *et al.*, 2010; Higgins *et al.*, 2009; Chu *et al.*, 2020) is given as

$$\tilde{\theta}_{\mathrm{PE}}(0.5) = \frac{n_2 + I(X = 0\,\mathrm{or}\,n)}{X_2 + 0.5I(X = 0\,\mathrm{or}\,n)}. \tag{10}$$

For ease of notation, we correspondingly denote this estimator as the piecewise Walter estimator in this paper.

(iv) To estimate RR, Carter et al. (2010) proposed another estimator as $\tilde{\mathrm{RR}}(1) = (X_1 + 1)(n_2 + 2)/[(X_2 + 1)(n_1 + 2)]$. Or equivalently, this yields the Carter estimator for the inverse proportion as

$$\tilde{\theta}(1) = \frac{n_2 + 2}{X_2 + 1}, \tag{11}$$

which is a special case of estimator (3) with $c = 1$.

2.2 The Horvitz–Thompson Estimator

On random sampling without replacement from a finite population, it is known that the Horvitz–Thompson estimator has played an important role in the literature for estimating the population total (Horvitz & Thompson., 1952; Cochran, 2007).

Let U be a population composed of t units $\{u_1, \ldots, u_t\}$, and p_i be the first-order selection probability associated with unit u_i. Let also $\boldsymbol{\Omega}$ be a random variable associated with the population U, and $\boldsymbol{\Omega}_i$ be the value of $\boldsymbol{\Omega}$ determined by unit u_i. Following these notations, the population total of $\boldsymbol{\Omega}$ can be defined as $T = \sum_{i=1}^{t} \boldsymbol{\Omega}_i$. Then as an unbiased estimator of T, the Horvitz–Thompson estimator is given as

$$\hat{T} = \sum_{j \in V} \omega_j \theta_j = \sum_{j \in V} \frac{\omega_j}{p_j}, \tag{12}$$

where ω_j is the observed value of $\boldsymbol{\Omega}_j$, and $V \subseteq \{1, \ldots, t\}$ is a subset of samples selected for estimating the population total. In practice, it is not uncommon that the inverse proportions $\theta_j = 1/p_j$ are unknown and so need to be estimated.

To estimate θ_j in (12), Fattorini (2006) proposed a numerical method via Monte Carlo simulations. Specifically in each simulation, a total of n samples were selected independently with replacement from the population U, with X_j being the number of samples that contain the j th unit, where $j \in V$. Further, to avoid the zero-event problem on X_j, Fattorini applied estimator (4) with $c = 1$ to estimate the inverse proportions by

$$\hat{\theta}_j(1) = \frac{n+1}{X_j + 1}, \quad j \in V. \tag{13}$$

This leads to the empirical Horvitz–Thompson estimator of the population total as $\hat{T}_m = \sum_{j \in V} \omega_j \hat{\theta}_j(1)$. Unless otherwise specified, we will ignore the subscript j in (13) and refer to $\hat{\theta}(1)$ as the Fattorini estimator.

For the Fattorini estimator in family (4) with $c = 1$, Seber (2013) showed that

$$E[\hat{\theta}(1)] = E\left(\frac{n+1}{X+1}\right) = \frac{1 - (1-p)^{n+1}}{p} = \theta - \theta\left(1 - \frac{1}{\theta}\right)^{n+1}. \tag{14}$$

Then by the fact that $\lim_{n\to\infty} \mathrm{Bias}[\hat{\theta}(1)] = \lim_{n\to\infty}[-\theta(1 - 1/\theta)^{n+1}] = 0$ for any fixed $\theta \in (1, \infty)$, the Fattorini estimator is an asymptotically unbiased estimator of θ when n is large. In addition, when p is large enough, or equivalently when θ is close to 1, the estimation bias of the Fattorini estimator is often negligible no matter whether n is large or small.

3 METHODOLOGY

3.1 Comparison of the Existing Estimators

In view of the demand for accurate estimation of the inverse proportion, we revisit the three families of shrinkage estimators in (3), (4) and (9) and compare them in both theory and practice. We first show that the three estimators are all consistent and asymptotically equivalent, with the proof of the theorem in Appendix A in the supporting information.

Theorem 1. *Let X be a binomial random variable with parameters n and p. For the shrinkage estimators in (3), (4) and (9) with any finite $c > 0$, we have the following properties:*

(i) $\tilde{\theta}(c)$, $\hat{\theta}(c)$ and $\tilde{\theta}_{\mathrm{PE}}(c)$ are all consistent estimators of θ;

(ii) $\tilde{\theta}(c)$, $\hat{\theta}(c)$ and $\tilde{\theta}_{\mathrm{PE}}(c)$ are all asymptotically equivalent such that $\sqrt{n}(\check{\theta} - \theta) \xrightarrow{D} N(0, \theta^2(\theta - 1))$, where $\check{\theta}$ is a generic notation for the three estimators and $\xrightarrow{D}$ denotes convergence in distribution.

Despite the asymptotic equivalence, we note however that their finite sample performance can be quite different. To demonstrate it, we conduct a numerical study to compare the three estimators under the Stein loss (SL), which, as a scale-invariant loss function, provides a better criterion for assessing the accuracy of θ compared with the location-invariant squared loss (Tong & Wang, 2007). It is noteworthy that, for comparison purpose, we have also presented the results for the squared loss in Figure S1. For the setting of parameters, we consider $\theta = 1.02$, 2 or 50, which is equivalent to $p = 0.98$, 0.5 or 0.02. We also consider $n = 10$ or 200 to represent the small and large sample sizes respectively, and let c range from 0 to 2 so as to cover most common choices of c in the literature. Then for each setting, we generate $N = 1{,}000{,}000$ data sets from the binomial distribution and estimate θ by each estimator from the three families. Finally, with the simulated data sets, we compute the Stein loss of each estimator by

$$\mathrm{SL}(\check{\theta}_k) = \frac{1}{N}\sum_{k=1}^{N}\left[\frac{\check{\theta}_k}{\theta} - \ln\left(\frac{\check{\theta}_k}{\theta}\right) - 1\right], \tag{15}$$

and then report the simulation results in Figure 1.

From Figure 1, it is evident that the estimators from family (4) perform better than those from the other two families in most settings. In particular, no estimator from family (3) is able to provide an accurate estimate when $\theta = 1.02$, no matter whether the sample size is large or small. On the other side, the estimators from family (9) fail to provide a stable performance when θ is moderate to large. To summarise, except for the extreme case where θ is relatively large and n is relatively small, the estimators from family (4) are always among the best and so can be safely recommended. Moreover, we also provide another evidence from the perspective of bias in Theorem 2 that the estimators from family (3) can be suboptimal for practical use, with the proof in Appendix B in the supporting information. Taken together, we will focus on the estimators $\hat{\theta}(c)$ in family (4) and propose to find the optimal c value that minimises the estimation bias.

Theorem 2. *Let X be a binomial random variable with parameters n and p. Then for the estimators from family (3), there does not exist a shrinkage parameter c such that $E[\tilde{\theta}(c)] = \theta$ when $p = 0.5$, or equivalently, when $\theta = 2$.*

3.2 Optimal Estimation of $\boldsymbol{\theta}$

For the estimators from family (4), we have introduced the Fattorini estimator with $c = 1$ as a special case with the asymptotic property in Section 2.2. However, as is shown in the numerical study, the Fattorini estimator may not provide an accurate estimate for the inverse proportion when n is small and θ is large. To further illustrate it, we take $n = 10$ and $\theta = 50$; then according to (14), the relative bias of the Fattorini estimator is as large as

$$\frac{E[\hat{\theta}(1)] - \theta}{\theta} \times 100\% = -(1 - 0.02)^{11} \times 100\% \approx -80.07\%.$$

In addition, it is noteworthy that the expected value of the Fattorini estimator is always lower than θ and so is consistently negatively biased. These evidences indicate that the Fattorini estimator may not be the optimal estimator in family (4).

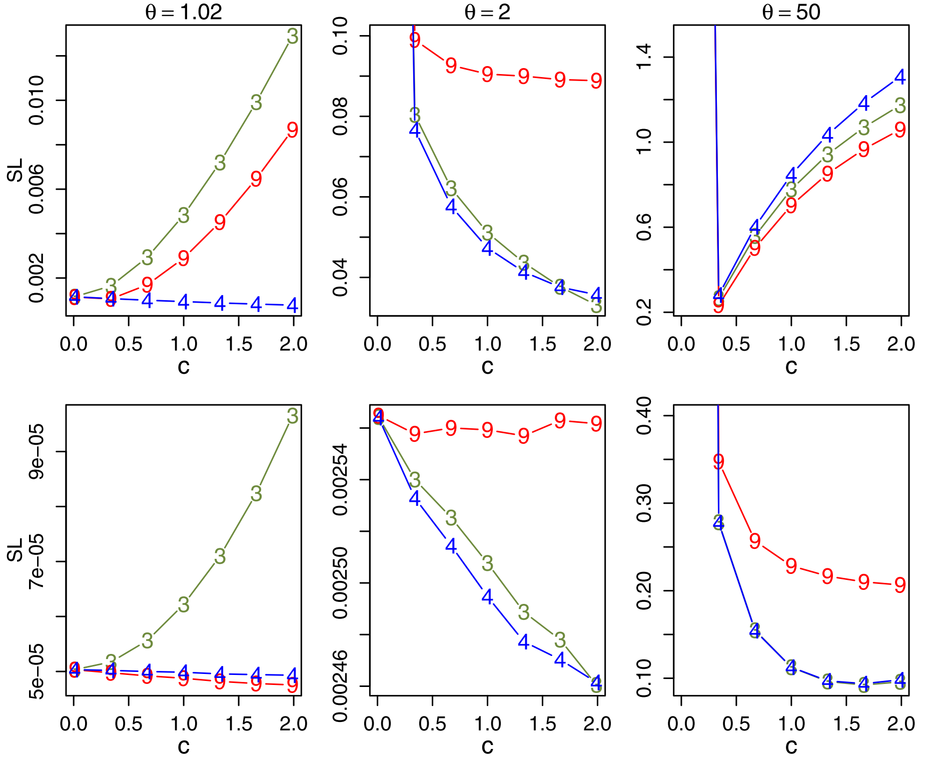

FIGURE 1. *The Stein losses for the shrinkage estimators from the three families with $\theta = 1.02$, 2 or 50, $n = 10$ (top three panels) or 200 (bottom three panels), and $c \in (0, 2)$, where '3' represents the estimators from family (3), '4' represents the estimators from family (4), and '9' represents the estimators from family (9).*

To eliminate the bias in the Fattorini estimator, we now define the optimal shrinkage parameter c as the value such that $E[\hat{\theta}(c)] = \theta$. For ease of notation, we also express the expected value of $\hat{\theta}(c)$ as

$$g(c) = E[\hat{\theta}(c)] = \sum_{x=0}^{n} \left(\frac{n+c}{x+c}\right) \binom{n}{x} p^x (1 - p)^{n-x}, \tag{16}$$

and then regard $g(c)$ as a function of c. In the following theorem, we provide some properties of $g(c)$, including the continuity, monotonicity and convexity, with the proof in Appendix C in the supporting information.

Theorem 3. *For the expected value function $g(c)$ in (16) with any finite integer n, we have the following properties:*

(i) $g(c)$ is a continuous function of c on $(0, \infty)$ with $\lim_{c \to 0} g(c) = \infty$ and $\lim_{c \to \infty} g(c) = 1$;
(ii) $g(c)$ is a strictly decreasing function of c on $(0, \infty)$;
(iii) $g(c)$ is a strictly convex function of c on $(0, \infty)$.

Note also that θ takes value on $(1, \infty)$, and $g(1) < \theta$ for any fixed n according to formula (14). Then by Theorem 3 and the Intermediate Value Theorem, there exists a unique solution $c \in (0, 1)$ such that $g(c) = \theta$, or equivalently,

$$g(c) = \sum_{x=0}^{n} \binom{n+c}{x+c} \binom{n}{x} p^x (1-p)^{n-x} = \frac{1}{p}. \tag{17}$$

When n is small, in particular for $n = 1$ or $n = 2$, we can derive the explicit solution of c from equation (17). When n is large, however, the degree of equation will be as high as $n + 1$, and consequently, an explicit solution for the unknown c may not exist. To summarise, we have the following theorem with the proof in Appendix D in the supporting information.

Theorem 4. *When n is less than 3, the solution of c in equation (17) is given by*

$$c_n = \begin{cases} p & n = 1, \\ p - 0.5 + \sqrt{0.5 - (p - 0.5)^2} & n = 2. \end{cases}$$

When $n \geq 3$, we have the approximate solution of c as

$$c_n \approx 1 - \frac{p^{-1}(1-p)^{n+1}}{(n+1)(1+D_1)D_2 - D_1}, \tag{18}$$

where

$$D_1 = \frac{1}{p(n+1)}\left[1 - (1-p)^{n+1}\right], \; D_2 = \frac{1}{p^2(n+1)(n+2)}\left[1 - (1-p)^{n+2} - (n+2)p(1-p)^{n+1}\right].$$

To check the accuracy of the approximate solution in Theorem 4, we also plot the numerical results of the true and approximate solutions of c as a function of p in Figure 2. Under various settings, we note that the true solution of c is given as a monotonically increasing function of p with the upper bound 1. And in addition, our approximate solution always works well as long as n or p is not extremely small.

3.3 Plug-In Estimator

To apply Theorem 4 for the optimal shrinkage parameter, we need a plug-in estimator for the unknown p. Intuitively, the MLE of p, $\hat{p}_{\text{MLE}} = X/n$, can serve as a natural choice. By doing so, however, for $n = 1$ we have $\hat{c}_1 = \hat{p}_{\text{MLE}} = X$, and further, it yields that $\hat{\theta}(\hat{c}_1) = (1 + \hat{c}_1)/(X + \hat{c}_1) = (1 + X)/2X$, which then suffers from the zero-event problem. For $n = 2$, it is noted that the same problem also remains. For $n \geq 3$, the approximate solution will no longer suffer from the zero-event problem; but on the other side, the denominator term, $(n+1)(1+D_1)D_2 - D_1$, in (18) will be zero when $X = n$, and consequently the approximate solution is still notapplicable. To conclude, the MLE of p cannot be directly applied as the plug-in estimator when applying Theorem 4 to estimate the inverse proportion.

To overcome the boundary problems on both sides, we consider the plug-in estimator of p with the following structure:

$$\tilde{p}_{\text{plug}}(\alpha) = \min(\max(\hat{p}_{\text{MLE}}, \alpha), 1 - \alpha),$$

where $0 < \alpha \leq 0.5$ is the threshold parameter. Then with $\tilde{p}_{\text{plug}}(\alpha)$ as the plug-in estimator of p, we let $\tilde{c}_n(\alpha)$ be the estimator of c_n in Theorem 4. To determine the best threshold value, we let θ

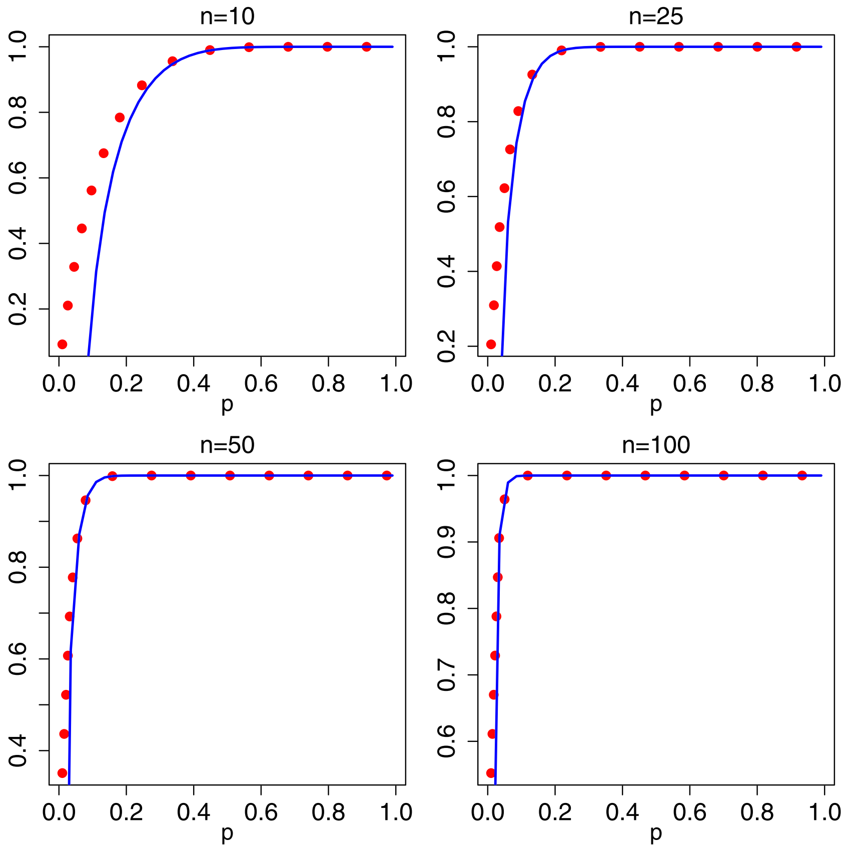

FIGURE 2. *The true and approximate solutions of c with n = 10, 25, 50 or 100. The solid dots represent the values of the true solution, and the solid lines represent the values of the approximate solution.*

range from 1.02 up to 50 and take several different combinations of n and α. Then with $N = 1{,}000{,}000$ data sets generated from the binomial distribution, we compute the relative bias of the estimator by

$$\mathrm{Bias}(\check{\theta}_k) = \frac{1}{N}\sum_{k=1}^{N}\left(\frac{\check{\theta}_k}{\theta} - 1\right), \tag{19}$$

where $\check{\theta}_k$ is a generic form of $\hat{\theta}_k(\tilde{c}_n(\alpha))$. Specifically in Figure 3, by taking $n = 1$, 2, 10 and 50, we plot the relative biases as functions of θ for our new estimator with $\alpha = 0.1$, 0.2, 0.3, 0.4, 0.5 and also for the Fattorini estimator.

From Figure 3, it is evident that $\alpha = 0.1$ may not provide an adequate remedy for the boundary problems. On the other side, when α tends to 0.5, our new estimator will perform more closely to the Fattorini estimator so that it may end up with an over-correction. Besides the relative bias, we have also presented the Stein loss of the estimators in Figure S2, which shows that the Stein loss is always an increasing function of α. Taken together, we conclude that $\alpha = 0.2$ is the minimum possible threshold that can provide an adequate correction. Finally, with the plug-in estimator $\tilde{p}_{\mathrm{plug}}(0.2) = \min(\max(\hat{p}_{\mathrm{MLE}}, 0.2), 0.8)$, our optimal estimator for the inverse proportion is given as

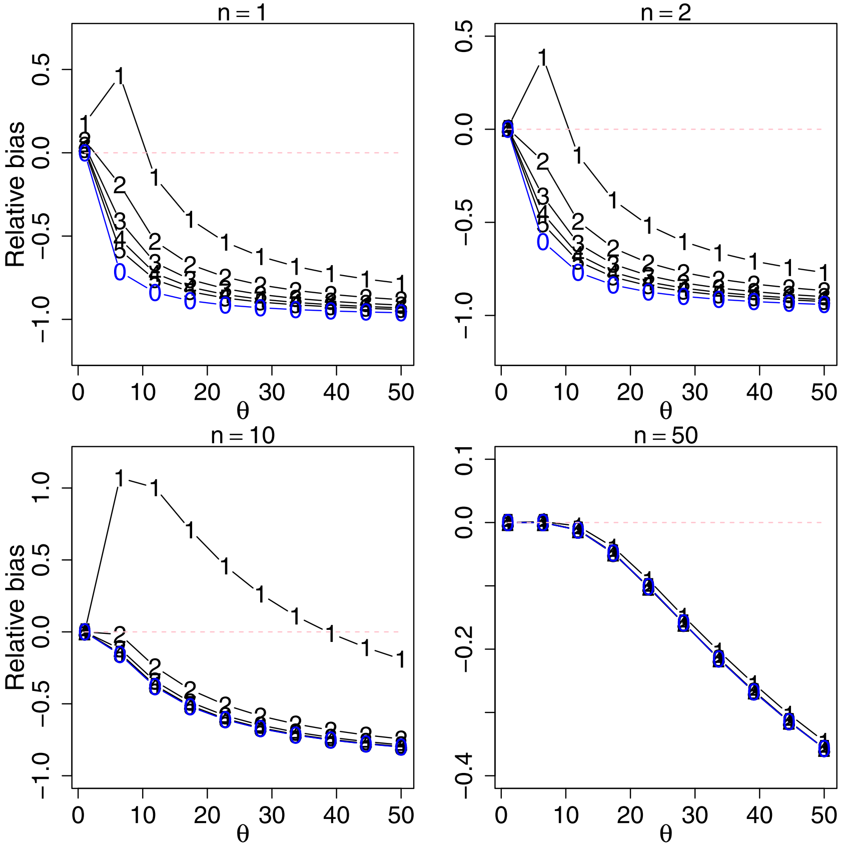

FIGURE 3. *The relative biases of $\hat{\theta}(\tilde{c}_n)$ with $\alpha = 0.1$, 0.2, 0.3, 0.4 or 0.5, where '1' represents the relative biases associated with $\alpha = 0.1$, '2' represents the relative biases associated with $\alpha = 0.2$, '3' represents the relative biases associated with $\alpha = 0.3$, '4' represents the relative biases associated with $\alpha = 0.4$, and '5' represents the relative biases associated with $\alpha = 0.5$. And for comparison, '0' represents the relative biases of the Fattorini estimator.*

$$\hat{\theta}(\tilde{c}_n) = \frac{n + \tilde{c}_n}{X + \tilde{c}_n}, \tag{20}$$

where $\tilde{c}_n = c_n(\tilde{p}_{\text{plug}}(0.2))$ is the estimator of c_n given in Theorem 4.

In the next theorem, we derive the asymptotic properties of estimator (20) with the proof in Appendix E in the supporting information. Note that the asymptotic variance of our new estimator is also $\theta^2(\theta - 1)/n$. In case an estimate of the asymptotic variance is needed, one can plug-in our new estimator in (20), which yields the variance estimate as $\hat{\theta}(\tilde{c}_n)^2[\hat{\theta}(\tilde{c}_n) - 1]/n$.

Theorem 5. *Let X be a binomial random variable with parameters n and p. For the estimator $\hat{\theta}(\tilde{c}_n)$ in (20), we have the following properties:*

(i) $\tilde{c}_n = 1 + o_p(1)$ and $\hat{\theta}(\tilde{c}_n)$ is a consistent estimator of θ;

(ii) $\hat{\theta}(\tilde{c}_n)$ is asymptotically equivalent to the estimators in (3), (4) and (9) such that $\sqrt{n}(\hat{\theta}(\tilde{c}_n) - \theta) \xrightarrow{D} N(0, \theta^2(\theta - 1))$.

4 Simulation Studies

In this section, we conduct simulation studies to evaluate the finite sample performance of our new estimator in (20) for the inverse proportion. For comparison, five existing estimators in the literature are also considered, including the Walter estimator in (7), the Pettigrew estimator in (8), the piecewise Walter estimator in (10), the Carter estimator in (11), and the Fattorini estimator in (13). For the simulation settings, we let θ range from 1.02 up to 50, which is equivalent to p ranging from 0.98 down to 0.02, and we consider $n = 1$, 2, 10, 50 or 200 as five different sample sizes. We further generate $N = 1{,}000{,}000$ data sets from the binomial distribution with each combination of θ and n. Finally, for the six estimators, we compute their relative biases by (19) and the Stein losses by (15) and then report them in Figures 4 and 5. For more comparison, the mean squared errors of the six estimators are also reported in Figure S3 with some discussion.

When $n = 1$ or 2, Figure 4 shows that the new estimator outperforms all other estimators in terms of the relative bias and Stein loss in most settings, as long as θ is not very small. In addition, it is evident that the Fattorini estimator may not provide an accurate estimate for θ when n

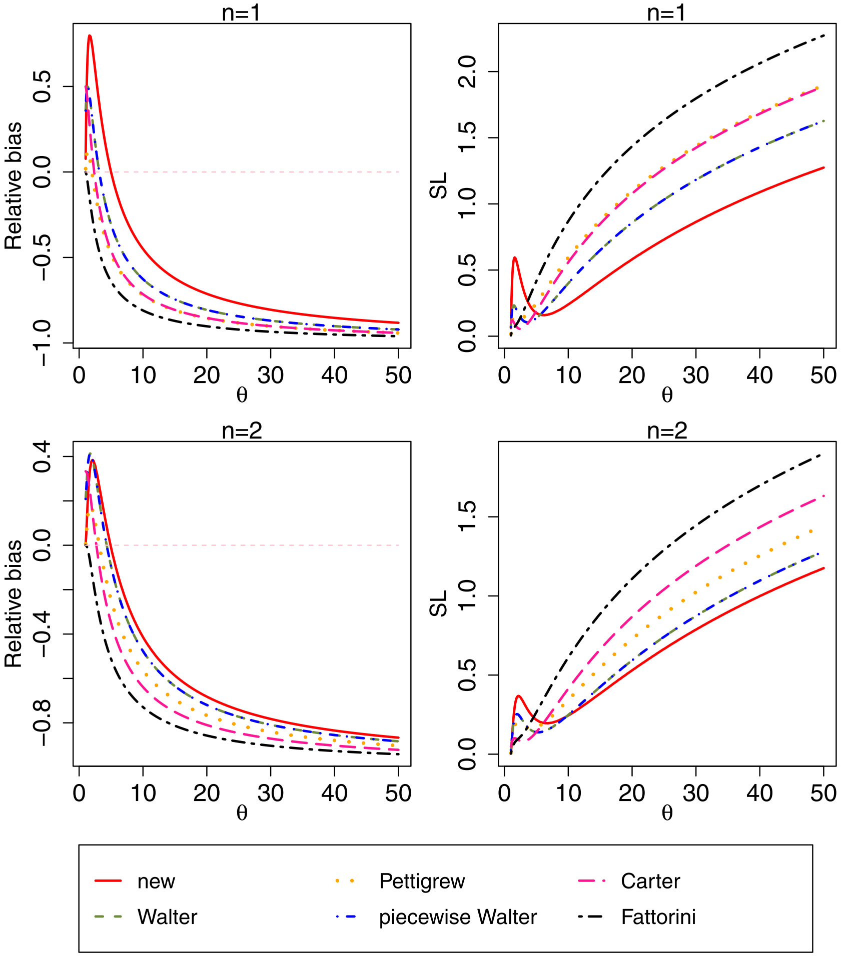

FIGURE 4. *The relative biases and the Stein losses of the six estimators with $n = 1$ or 2.*

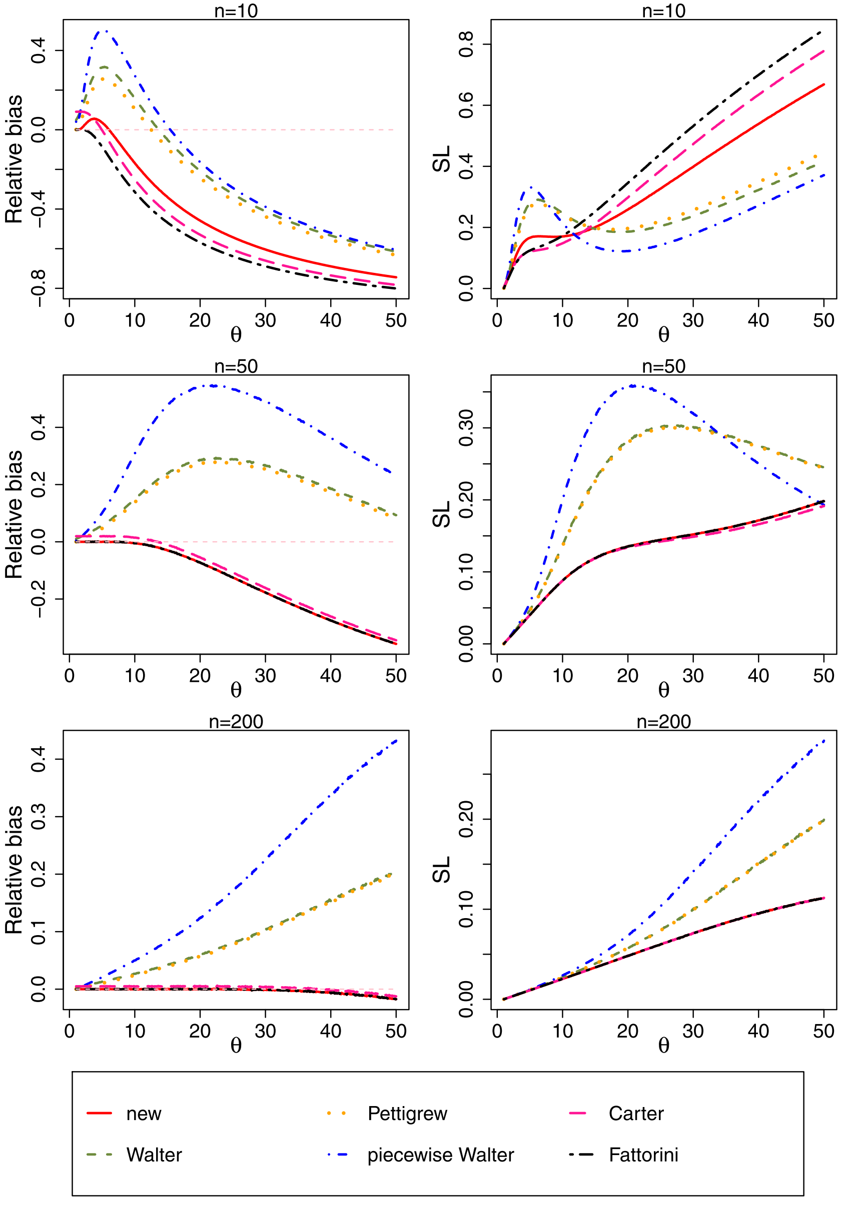

FIGURE 5. *The relative biases and the Stein losses of the six estimators with $n = 10$, 50 or 200.*

is extremely small. From Figure 5, we observe that the new estimator performs better than the Carter and Fattorini estimators when the sample size is moderate ($n = 10$), or performs comparably to them when the sample size is large ($n = 50$ or 200). In contrast, the other three estimators, including the Walter, Pettigrew and piecewise Walter estimators, belong to another league. Specifically, they perform well when the sample size is moderate, but fail to provide a stable performance when the sample size is large.

To conclude, our new estimator performs best in most settings when the sample size is small to moderate, and performs as well as other estimators when the sample size is large. For ease of implementation, we have also provided the R code for the new estimator in Appendix G in the supporting information. Finally, as a price to pay, we note that our plug-in estimator in (20) has a more complex form than the existing estimators, even though it does not increase much computational cost (for details, see Appendix G in the supporting information). In view of this, we also highly recommend to use the Carter and Fattorini estimators by virtue of their simple forms and the good performance when the sample size is large.

5 An Application to Zero-Event Studies

In this section, we apply our new estimator to a meta-analysis on COVID-19 data with zero-event studies. Chu *et al.* (2020) carried out an excellent review to investigate effects of physical distancing, face masks and eye protection on the infection of severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This systematic review was published in June 2020 and is now attracting more and more attention. For example, in Google Scholar as of 14 March 2023, their paper has received a total of 4,170 citations. Also as commented by MacIntyre & Wang (2020), this systematic review provides a landmark for people to be aware of the

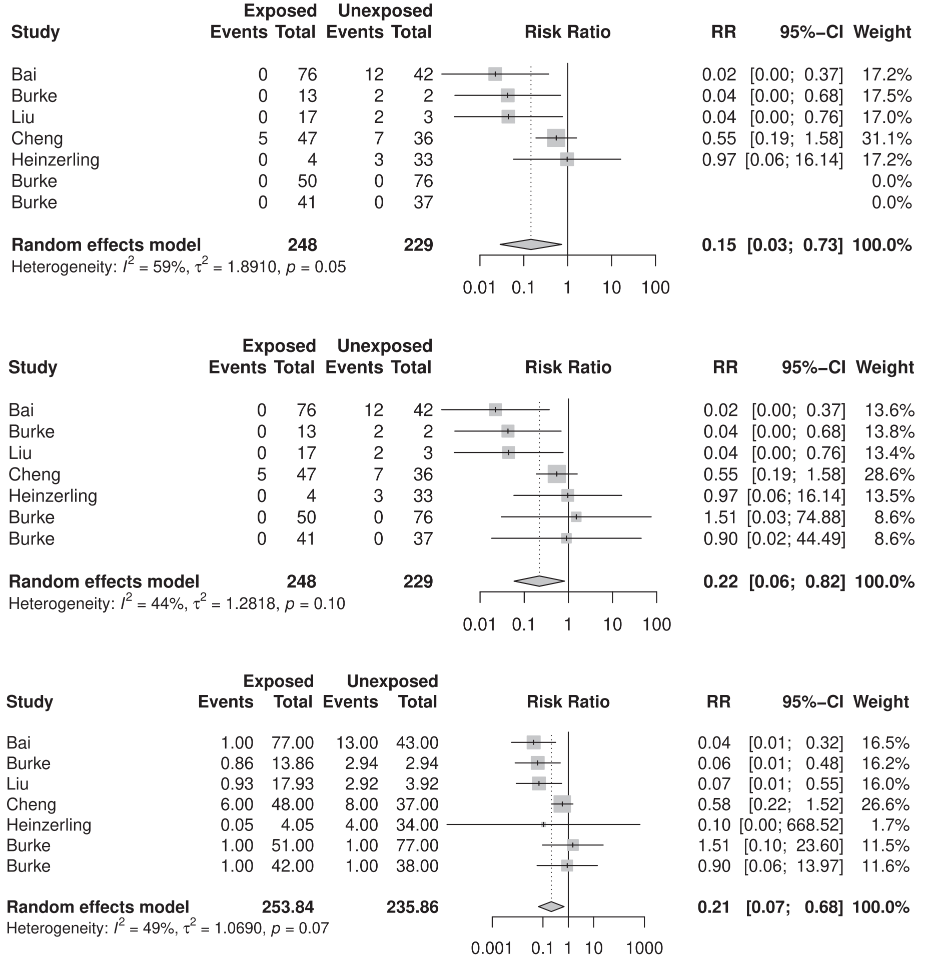

FIGURE 6. *Forest plots on the relative risk between physical distancing and infection for COVID-19 data.*

importance of physical distancing and face protection. In particular for physical distancing, they applied the relative risks as effect sizes and concluded that the virus transmission is significantly reduced with a further distance.

In the top panel of Figure 6, seven studies were included in their meta-analysis of physical distancing for COVID-19 data, where six studies therein suffered from the zero-event problem. For the four single-zero-event studies, the 0.5 continuity correction was added to all the counts of events. For the two double-zero-event studies, they were not included in the meta-analysis. By Xu *et al.* (2020) and our simulation results, adding the 0.5 continuity correction is suboptimal. Moreover, Xu *et al.* (2020) also showed that the double-zero-event studies may also be informative, and so excluding them can be questionable and/or even alter the results. In view of the above limitations, we re-conducted the meta-analysis on COVID-19 data that also includes the two double-zero-event studies. Specifically, by applying our new estimator in (20), the relative risks are estimated by

$$\hat{\mathrm{RR}}(\tilde{c}_n) = \frac{(X_1 + \tilde{c}_{n_1})(n_2 + \tilde{c}_{n_2})}{(X_2 + \tilde{c}_{n_2})(n_1 + \tilde{c}_{n_1})}, \tag{21}$$

where $\tilde{c}_{n_1}$ and $\tilde{c}_{n_2}$ are the estimates of the optimal shrinkage parameter for the exposed group and the unexposed group, respectively. For comparison, we also conduct a meta-analysis for all seven studies by the 0.5 continuity correction, and then present all the forest plots in Figure 6.

From the middle and bottom panels of Figure 6, it is evident that the new meta-analytical results with the double-zero-event studies also support the claim that a further distance will reduce the virus infection. On the other hand, the evidence becomes less significant when the combined relative risks get larger. Moreover, by comparing the two forest plots that both contain the double-zero-event studies, we note that their combined relative risks are also close, whereas our new estimator in (21) yields a narrower confidence interval.

6 Discussion

In this paper, we first reviewed the existing estimators for the inverse proportion, or formally the reciprocal of a binomial proportion. We then proposed a new estimator of the inverse proportion by deriving the optimal shrinkage parameter c in the family of estimators (4). Simulation studies showed that our new estimator performs better than, or as well as, the existing competitors in most settings. Finally, we also applied our new estimator to a recent meta-analysis on COVID-19 data with the zero-event problem, and our findings provided some additional evidence for addressing the scientific question: 'how does physical distancing effectively prevent the transmission of the new coronavirus?'

To highlight the main contributions of this paper, we have made a good effort in finding the optimal estimator for the inverse proportion related to the binomial distribution. According to Gupta (1967), there does not exist an unbiased estimator for the inverse proportion θ. To verify this result, by the proof-by-contradiction we assume that $\hat{\theta}_{\mathrm{u}} = \eta(X)$ is an unbiased estimator of θ. Then by definition, $E(\hat{\theta}_{\mathrm{u}}) = \sum_{x=0}^{n} \eta(x) \binom{n}{x} p^x (1 - p)^{n-x} = \theta$. From the left-hand side, the expected value of $\hat{\theta}_{\mathrm{u}}$ is a polynomial of p with degree n. For the right-hand side, by the Taylor expansion we have $\theta = 1/p = \sum_{i=0}^{\infty} (1 - p)^i$, which is a polynomial of p with infinite degree. This shows that the unbiasedness cannot be held for any finite n. In view of this property, there is probably no uniformly best estimator for the inverse proportion. Although we have conducted some nice work in this paper, we believe that more advanced research is still needed to further improve the estimation accuracy of the inverse proportion. For example, one may develop a better and more robust approximation for the optimal shrinkage parameter when the binomial proportion p is extremely small. In addition, other families of shrinkage estimators can also be considered to see whether they can yield better estimators for the inverse proportion.

Last but not least, we note that the new estimation of the inverse proportion can have many other real applications. For instance, the spirit of our new method may also be applied to estimate the number needed to treat (NNT), which is another important medical term and was first introduced by Laupacis *et al.* (1988). Specifically, NNT is defined as $\mathrm{NNT} = 1/(p_1 - p_2)$, where p_1 is the event probability in the exposed group and p_2 is the event probability in the unexposed group. Noting also that $p_1 - p_2$ is the absolute risk reduction (ARR), NNT can be explained as the average number of patients who are needed to be treated to obtain one more patient cured compared with a control in a clinical trial (Hutton, 2000). Nevertheless, the estimation of NNT will be more challenging than the estimation of the inverse proportion, mainly because the estimate of $p_1 - p_2$ can be either positive or negative, in addition to the zero-event problem in the denominator. More recently, Veroniki *et al.* (2019) also referred to this situation as the statistically nonsignificant result, which may lead to an unexpected calculation complication.

In addition to NNT, Zhang & Yin (2021) proposed the reduction in number to treat (RNT) as a new measure of the treatment effect in randomised control trials. Specifically, let the two inverse proportions $\theta_1 = 1/p_1$ be the average number of patients who are needed to be treated to obtain one patient cured in the exposed group and $\theta_2 = 1/p_2$ be the average number of patients who are needed to be treated to obtain one patient cured in the unexposed group, then RNT is defined as $\mathrm{RNT} = \theta_2 - \theta_1 = 1/p_2 - 1/p_1$. Also by (2), the MLE of RNT is given as $\hat{\mathrm{RNT}}_{\mathrm{MLE}} = n_2/X_2 - n_1/X_1$, which once again may not be applicable when the value of X_1 or X_2 is zero. Thus to study the statistical inference of RNT, it also requires a valid estimate for each of the inverse proportions that does not suffer from the zero-event problem. We expect that our new work in this paper will shed light on new directions on the NNT and RNT estimation, which can be particularly useful in clinical trials and evidence-based medicine.

SUPPORTING INFORMATION

The readers may refer to the Supporting Information for the technical details, the additional simulations with figures, and the R code for implementing the new estimator.

ACKNOWLEDGEMENTS

The authors thank the Co-Editor-in-Chief, an Editor, and two anonymous referees for their constructive comments that have led to a significant improvement of this paper. Ping He's research was supported by the Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College (2022B1212010006). Tiejun Tong's research was supported by the General Research Funds (HKBU12303918 and HKBU12303421), the Initiation Grant for Faculty Niche Research Areas (RC-FNRA-IG/20-21/SCI/03) of Hong Kong Baptist University, and the National Natural Science Foundation of China (1207010822).

REFERENCES

Agresti, A. (2003). *Categorical data analysis*, 2. John Wiley & Sons: Hoboken.

Agresti, A. & Coull, B.A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *Amer. Statist.*, **52**, 119–126.

Altman, D.G. (1998). Confidence intervals for the number needed to treat. *Br. Med. J.*, **317**, 1309–1312.

Brown, L.D., Cai, T.T. & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statist. Sci.*, **16**, 101–117.

Carter, R.E., Lin, Y., Lipsitz, S.R., Newcombe, R.G. & Hermayer, K.L. (2010). Relative risk estimated from the ratio of two median unbiased estimates. *J. Roy. Statist. Soc. Ser. C*, **59**, 657–671.

Chao, M.T. & Strawderman, W.E. (1972). Negative moments of positive random variables. *J. Amer. Statist. Assoc.*, **67**, 429–431.

Chu, D.K., Akl, E.A., Duda, S., Solo, K., Yaacoub, S., Schünemann, H.J. & COVID-19 Systematic Urgent Review Group Effort (SURGE) study authors (2020). Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: A systematic review and meta-analysis. *Lancet*, **395**, 1973–1987.

Cochran, W.G. (2007). *Sampling techniques*. John Wiley & Sons: New York.

Fattorini, L. (2006). Applying the Horvitz-Thompson criterion in complex designs: A computer-intensive perspective for estimating inclusion probabilities. *Biometrika*, **93**, 269–278.

Gamrot, W. (2013). On a class of estimators for a reciprocal of Bernoulli parameter. *Stud. Ekon.*, **133**, 71–85.

Good, I.J. 1980. Some history of the hierarchical Bayesian methodology. In *Bayesian statistics*, Eds. J. M. Bernardo, M. H. Degroot, D. V. Lindley & A. F. M. Smith, pp. 489-504. Valencia: University Press.

Gupta, M.K. (1967). Unbiased estimate for $1/p$. *Ann. Inst. Statist. Math.*, **19**, 413–416.

Hanley, J.A. & Lippman-Hand, A. (1983). If nothing goes wrong, is everything all right? Interpreting zero numerators. *J. Am. Med. Assoc.*, **249**, 1743–1745.

Hartung, J. & Knapp, G. (2001). A refined method for the meta–analysis of controlled clinical trials with binary outcome. *Stat. Med.*, **20**, 3875–3889.

Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J. & Welch, V.A. (2009). *Cochrane handbook for systematic reviews of interventions*, 2. John Wiley & Sons: Chichester.

Hogg, R.V., McKean, J.W. & Craig, A.T. (2005). *Introduction to mathematical statistics*. Pearson Education: Boston.

Horvitz, D.G. & Thompson., D.J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, **47**, 663–685.

Hutton, J.L. (2000). Number needed to treat: properties and problems. *J. Roy. Statist. Soc. Ser. A*, **163**, 381–402.

Laupacis, A., Sackett, D.L. & Roberts., R.S. (1988). An assessment of clinically useful measures of the consequences of treatment. *N. Engl. J. Med.*, **318**, 1728–1733.

MacIntyre, C.R. & Wang, Q. (2020). Physical distancing, face masks, and eye protection for prevention of COVID-19. *Lancet*, **395**, 1950–1951.

Pettigrew, H.M., Gart, J.J. & Thomas, D.G. (1986). The bias and higher cumulants of the logarithm of a binomial variate. *Biometrika*, **73**, 425–435.

Schwarzer, G. (2007). meta: An R package for meta-analysis. *R News*, **7**, 40–45.

Seber, G.A.F. (2013). *Statistical models for proportions and probabilities*. Springer: Heidelberg.

Tong, T. & Wang, Y. (2007). Optimal shrinkage estimation of variances with applications to microarray data analysis. *J. Amer. Statist. Assoc.*, **102**, 113–122.

Tuyl, F., Gerlach, R. & Mengersen, K. (2009). The rule of three, its variants and extensions. *Int. Stat. Rev.*, **77**, 266–275.

Veroniki, A.A., Bender, R., Glasziou, P., Straus, S.E. & Tricco, A.C. (2019). The number needed to treat in pairwise and network meta-analysis and its graphical representation. *J. Clin. Epidemiol.*, **111**, 11–22.

Walter, S.D. (1975). The distribution of Levin's measure of attributable risk. *Biometrika*, **62**, 371–372.

Wei, J., Lin, E., Shi, J., Yang, K., Hu, Z., Zeng, X. & Tong, T. (2021). Meta-analysis with zero-event studies: A comparative study with application to COVID-19 data. *Mil. Med. Res.*, **8**, 41.

Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference. *J. Amer. Statist. Assoc.*, **22**, 209–212.

Xu, C., Li, L., Lin, L., Chu, H., Thabane, L., Zou, K. & Sun, X. (2020). Exclusion of studies with no events in both arms in meta-analysis impacted the conclusions. *J. Clin. Epidemiol.*, **123**, 91–99.

Zhang, C. & Yin, G. (2021). Reduction in number to treat versus number needed to treat. *BMC Med. Res. Methodol.*, **21**, 48.

[*Received August 2021; accepted March 2023*]