RESEARCH ARTICLE

Statistics in Medicine WILEY

Ratio estimators of intervention effects on event rates in cluster randomized trials

Xiangmei Ma[1] | Paul Milligan[2] | Kwok Fai Lam[1,3] | Yin Bun Cheung[1,4,5]

[1]Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore, Singapore

[2]Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK

[3]Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong, China

[4]Programme in Health Services & Systems Research, Duke-NUS Medical School, Singapore, Singapore

[5]Tampere Center for Child, Adolescent and Maternal Health Research, Tampere University, Tampere, Finland

Correspondence

Yin Bun Cheung, Centre for Quantitative Medicine, Duke-NUS Medical School, 8 College Road, Singapore 169857, Singapore.
Email: yinbun.cheung@duke-nus.edu.sg

Funding information

National Medical Research Council, Grant/Award Number: MOH-000526

Abstract

We consider five asymptotically unbiased estimators of intervention effects on event rates in non-matched and matched-pair cluster randomized trials, including ratio of mean counts (r_1), ratio of mean cluster-level event rates (r_2), ratio of event rates (r_3), double ratio of counts (r_4), and double ratio of event rates (r_5). In the absence of an indirect effect, they all estimate the direct effect of the intervention. Otherwise, r_1, r_2, and r_3 estimate the total effect, which comprises the direct and indirect effects, whereas r_4 and r_5 estimate the direct effect only. We derive the conditions under which each estimator is more precise or powerful than its alternatives. To control bias in studies with a small number of clusters, we propose a set of approximately unbiased estimators. We evaluate their properties by simulation and apply the methods to a trial of seasonal malaria chemoprevention. The approximately unbiased estimators are practically unbiased and their confidence intervals usually have coverage probability close to the nominal level; the asymptotically unbiased estimators perform well when the number of clusters is approximately 32 or more per trial arm. Despite its simplicity, r_1 performs comparably with r_2 and r_3 in trials with a large but realistic number of clusters. When the variability of baseline event rate is large and there is no indirect effect, r_4 and r_5 tend to offer higher power than r_1, r_2, and r_3. We discuss the implications of these findings to the planning and analysis of cluster randomized trials.

KEYWORDS

cluster randomized trial, event rate, incidence rate ratio, ratio estimator, relative incidence

1 | INTRODUCTION

The cluster randomized trial (CRT) is an important study design in medical and health research.[1,2,3] Data on outcome events may be collected by passive surveillance or active surveillance.[4] Passive surveillance methods may or may not provide data at the individual level. That is, they may determine only the number of events in a cluster, without identifying

which individual members of the cluster experienced the events. Furthermore, the denominators for standard practice of calculating event rates may not be available.[4] The advantage of passive surveillance is that the monetary and opportunity cost of data collection can be much reduced.

Broadly speaking, there are two approaches to the analysis of CRTs: individual-level analysis and cluster-level analysis. Methods for individual-level analysis of CRTs include random-effects models and generalized estimating equations. As compared to cluster-level analysis, individual-level analysis has the relative advantage of efficiency and ease in covariate adjustment. However, it has the relative disadvantage of being less robust, especially when the number of clusters is small.[1] Furthermore, data collection by passive surveillance may not be compatible with individual-level analysis. In this manuscript we consider only cluster-level analysis.

An estimator of the intervention effect in terms of incidence rate ratio, also called relative incidence, that only uses event data is a ratio of the arithmetic mean of the number of outcome events per cluster in the intervention arm to that in the control arm. We call this the "ratio of mean counts", denoted by r_1.

The denominator of event rates, that is, units of person-time, in CRTs is usually variable across clusters. The person-time for estimation of an event rate is sometimes approximated by the population size at some point of the study duration. In this article we use the phrases person-time and population size interchangeably. Typical statistical practice makes comparison of event rates instead of mean number of events between trial arms. It requires extra resources in the collection of person-time data. A demographic surveillance system, a population census, or rounds of community surveys may be required for this purpose. With both the number of events and person-time collected for each cluster, one may calculate a cluster-level event rate for each cluster, denoted by $c_{ij} = y_{ij}/p_{ij}$, where y_{ij} and p_{ij} are the number of events and person-time in the jth cluster in the ith trial arm, respectively. Then, the arithmetic means of the cluster-level event rates in the intervention and control arms are calculated, denoted by $\bar{c}_1$ and $\bar{c}_0$, respectively. The ratio of the two means, $r_2 = \bar{c}_1/\bar{c}_0$ is a popular estimator of the incidence rate ratio.[1,5,6] This estimator has been evaluated by simulations but not analytically. One simulation study considered scenarios of, approximately, CV $(c_{ij}) = 0$, 0.125, and 0.25, where CV (c_{ij}) is the coefficient of variation of the cluster-level event rate in the ith trial arm and CV (c_{0j}) is known as "k" in the literature.[5] It found little bias in r_2. However, another simulation study considered a broader range of CV (c_{ij}).[6] It showed that r_2 was practically unbiased when CV $(c_{ij}) = 0.05$ and 0.15, but it was biased when CV $(c_{ij}) = 0.4$. Analytical investigation and simulation evaluation in a broad range of parameter values are warranted.

An alternative estimator of incidence rate ratio can be obtained by first calculating the event rate in each trial arm as the sum of the number of events divided by the sum of person-time over the clusters,[7,8] and then calculate the ratio of these event rate estimates between the trial arms. We call this the "ratio of event rates", denoted by r_3. While $\bar{c}_i$ ($i = 0,1$) is an unweighted average of cluster-level event rates in the ith trial arm, the alternative estimator of event rate here can be seen as a weighted average of cluster-level event rates, with the clusters' population sizes as weights.

In CRTs, interventions are often provided only to a specific group of the cluster members instead of all cluster members. For example, in studies of vaccines for pediatric infectious diseases, usually only young children in a specific age range are offered the interventions or its control comparators. Older children and adults are not. We refer to the two groups of cluster members as the target and non-target groups. The outcome events may occur in both groups. In studies based on passive surveillance, the event data may be collected for the non-target group in addition to the target group without much additional resources required, because the capital cost and infrastructure are already invested for the target group anyway. We consider an estimator that we call "double ratio of counts", denoted by r_4, by replacing the sums of person-time in r_3 by the sums of number of events in the non-target groups. Note that this estimator is defined even if the number of events in the non-target group is zero in some clusters, which is a realistic situation because usually the reason of it being a non-target group is that the disease incidence is relatively low. The motivation for considering this estimator arises from not only concerns of feasibility and cost of data collection but also concerns of precision and power. From an epidemiological point of view, sometimes we anticipate that the event counts in the target and non-target groups are highly correlated, because they are both the manifestation of the disease burden in the clusters. In particular, some events are highly localized, for example, infectious diseases occurring in small outbreaks. For such events, the correlation between number of events in the target and non-target groups, say children inside and outside a vaccination age range, is likely to be much stronger than the correlation between the number of events and amount of person-time in the target group. This advantage in correlation offers a potential for improved precision. Note that r_4 and the three estimators aforementioned have different targets of estimation (estimands): r_1, r_2, and r_3 estimate the total effect of the intervention whereas r_4 estimates the direct effect. Details will be discussed in the next section.

We also propose a new estimator that we call "double ratio of event rates", denoted by r_5. It has the ratio of event rates between the target and non-target groups in the intervention arm as the numerator and its counterpart in the control arm as the denominator. Details in statistical notations will be provided in the next section. We hypothesize that this estimator will out-perform r_4 in precision and power.

Donner and Klar pointed out that CRTs of binary outcomes may regard a proportion as a ratio and then an appropriate variance estimate can be obtained from sample survey theory.[9] They used the ratio of this estimated variance to the estimated binomial variance to adjust the Chi-square statistics for hypothesis testing. In the context of toxicological experiments in which litters of animals were the experimental units and a binary outcome was observed for each animal, Rao and Scott proposed using the aforementioned approach to adjust the Chi-square and Cochran-Armitage statistics.[10] There has been some subsequent research on using ratio estimators for CRTs with event rate outcomes, including the two simulation studies of r_2 aforementioned.[5,6] Furthermore, Dufault and Jewell proposed permutation tests of counts of events only, with or without adjustment for differential ascertainment.[4] All of them concerned only CRTs that randomize clusters individually and (implicitly) aim to estimate the total effect.

This study aims to (a) evaluate and compare the performance of the five estimators aforementioned and (b) develop, evaluate, and compare bias-corrected version of them. In Section 2 we will analytically assess and develop the methods. In Section 3 we will evaluate the methods by simulation in a broad range of realistic scenarios. In Section 4 we will apply the methods to a study of seasonal malaria chemoprevention. Section 5 gives some concluding remarks.

For brevity, we will focus on CRTs that randomize clusters individually, that is, non-matched CRTs. Where necessary we also provide the details for matched-pair CRTs in which one cluster per matched pair is randomized to receive the intervention and the other serves as the control. Introduction to the two types of CRTs can be found in, for example, Hayes and Moulton[1] and Donner and Klar.[2]

2 | STATISTICAL METHODS

2.1 | Intervention effects and event rates

An intervention may have a direct effect and an indirect effect, for example, via reducing disease transmission in the community.[3,11] Only the intervention's target group can benefit from the direct effect; both the target and non-target groups may benefit from the indirect effect, if any. Assume that:

$$\beta_T = \beta_D + \beta_I$$

where β_T, β_D, and β_I are the total, direct, and indirect effects in terms of log incidence rate ratio. If there is no indirect effect, $\beta_I = 0$ and $\beta_T = \beta_D$. The presence of an indirect effect depends on various factors including the nature of the interventions and outcome events. For example, even though vaccines are often anticipated to generate some degree of indirect effect on efficacy endpoints, they are usually anticipated to have no indirect effect on safety endpoints.

Let Y_{ijk} and P_{ijk} be the number of events and total person-time in the kth group of the jth cluster in the ith trial arm in the population, respectively, where $k = 1$ and 0 represent the target and non-target groups, respectively, $i = 1$ and 0 represent intervention and control trial arms, respectively. We consider a data generating process that is often used in epidemiologic modeling, that the expected value of Y_{ijk} given P_{ijk} is:

$$E\left(Y_{ijk}|P_{ijk}\right) = \exp\left(\beta_{D,ik} + \beta_{I,i} + \alpha_{ij} + \gamma_k\right) P_{ijk}, \tag{1}$$

where $\beta_{D,ik}$ and $\beta_{I,i}$ represent direct and indirect effects ($\beta_{D,11} = \beta_D$; $\beta_{I,1} = \beta_I$; $\beta_{D,10} = \beta_{D,01} = \beta_{D,00} = \beta_{I,0} = 0$), α_{ij} is a random cluster effect with standard deviation SD $\left(\alpha_{ij}\right) \geq 0$ that represents variation in event rates between clusters within each trial arm, and γ_k represents the difference in event rates between the target and non-target group ($\gamma_1 = \gamma$; $\gamma_0 = 0$). Note that SD $\left(\alpha_{ij}\right) \cong$ CV $\left(c_{ij}\right)$, the coefficient of variation in cluster-level event rate.[5,6] By randomization, the distributions of α_{ij} and P_{ijk} are identical in expectation between the intervention and control arms. From Equation (1), CV $\left(Y_{ijk}\right) =$ CV $\left(P_{ijk}\right)$ and corr $\left(Y_{ijk}, P_{ijk}\right) = 1$ if SD $\left(\alpha_{ij}\right) = 0$. The difference CV $\left(Y_{ijk}\right) -$ CV $\left(P_{ijk}\right)$ increases and the correlation corr $\left(Y_{ijk}, P_{ijk}\right)$ decreases as SD $\left(\alpha_{ij}\right)$ increases.

2.2 | Asymptotically unbiased estimators

Given a sample dataset of $\{y_{ijk}, p_{ijk} : j = 1, 2, \ldots, n_i; i = 0, 1; k = 0, 1\}$, with randomization and a large number of clusters per trial arm, the ratio of mean counts, r_1, provides an asymptotically unbiased estimator of the total effect that compares the event rates in the intervention and control arms:

$$r_1 = \frac{\bar{y}_1}{\bar{y}_0} = \frac{\sum_{j=1}^{n_1} y_{1j1}/n_1}{\sum_{j=1}^{n_0} y_{0j1}/n_0} \approx \frac{\sum_{j=1}^{n_1} \exp(\beta_D + \beta_I + \alpha_{1j} + \gamma) p_{1j1}/n_1}{\sum_{j=1}^{n_0} \exp(\alpha_{0j} + \gamma) p_{0j1}/n_0} = \exp(\beta_D + \beta_I) \frac{\sum_{j=1}^{n_1} \exp(\alpha_{1j} + \gamma) p_{1j1}/n_1}{\sum_{j=1}^{n_0} \exp(\alpha_{0j} + \gamma) p_{0j1}/n_0} \approx \exp(\beta_T).$$

When n_1 and n_0 are small, $\left[\sum_{j=1}^{n_1} \exp(\alpha_{1j} + \gamma) p_{1j1}/n_1\right] / \left[\sum_{j=1}^{n_0} \exp(\alpha_{0j} + \gamma) p_{0j1}/n_0\right] \not\approx 1$,

causing a small sample bias in the ratio estimator.[7] This and the bias in the other estimators will be discussed in Section 2.3.

Similarly, the ratio of mean cluster-level event rates,

$$r_2 = \frac{\bar{c}_1}{\bar{c}_0} = \frac{\sum_{j=1}^{n_1} c_{1j}/n_1}{\sum_{j=1}^{n_0} c_{0j}/n_0},$$

where $c_{ij} = y_{ij1}/p_{ij1}$, and ratio of event rates,

$$r_3 = \frac{R_1}{R_0} = \frac{\sum_{j=1}^{n_1} y_{1j1}/\sum_{j=1}^{n_1} p_{1j1}}{\sum_{j=1}^{n_0} y_{0j1}/\sum_{j=1}^{n_0} p_{0j1}},$$

also provide asymptotically unbiased estimators of the total effect.

The variance of r_1, r_2, and r_3 can be written as:[5,7,8]

$$\text{Var}(r_1) = r_1^2 \left[\frac{\text{CV}(y_{1j1})^2}{n_1} + \frac{\text{CV}(y_{0j1})^2}{n_0}\right], \tag{2}$$

$$\text{Var}(r_2) = r_2^2 \left[\frac{\text{Var}(\bar{c}_1)}{\bar{c}_1^2} + \frac{\text{Var}(\bar{c}_0)}{\bar{c}_0^2}\right], \tag{3}$$

and

$$\text{Var}(r_3) = r_3^2 \left[\frac{\text{Var}(R_1)}{R_1^2} + \frac{\text{Var}(R_0)}{R_0^2}\right]. \tag{4}$$

Furthermore, in first-order Taylor series expansion, $\text{Var}(\bar{c}_i) = \text{Var}(R_i)$,[8] and

$$\text{Var}(R_i) = \frac{R_i^2}{n_i}\left[\text{CV}(y_{ij1})^2 + \text{CV}(p_{ij1})^2 - 2\text{CV}(y_{ij1})\,\text{CV}(p_{ij1})\,\text{corr}(y_{ij1}, p_{ij1})\right].$$

From Equation (1), $\text{CV}(y_{ijk}) \geq \text{CV}(p_{ijk})$. Let $\text{CV}(p_{ijk}) = \theta_i \times \text{CV}(y_{ijk})$, where $0 < \theta_i \leq 1$.

If r_1 and r_2 give the same sample estimate, a comparison of $\text{Var}(r_1)$ and $\text{Var}(r_2)$ boils down to an evaluation of whether

$$n_i\left[\frac{\text{Var}(\bar{c}_i)}{\bar{c}_i^2} - \frac{\text{CV}(y_{ij1})^2}{n_i}\right] \approx n_i\left[\frac{\text{Var}(R_i)}{\bar{c}_i^2} - \frac{\text{CV}(y_{ij1})^2}{n_i}\right]$$

$$= \left(\frac{R_i}{\bar{c}_i}\right)^2\left[\text{CV}(y_{ij1})^2 + \text{CV}(p_{ij1})^2 - 2\text{CV}(y_{ij1})\,\text{CV}(p_{ij1})\,\text{corr}(y_{ij1}, p_{ij1})\right] - \text{CV}(y_{ij1})^2$$

$$= \left(\frac{R_i}{\bar{c}_i}\right)^2\left[(1 + \theta_i^2 - 2\theta_i\text{corr}(y_{ij1}, p_{ij1}))\,\text{CV}(y_{ij1})^2\right] - \text{CV}(y_{ij1})^2 < 0.$$

Therefore, $\text{Var}(r_2) < \text{Var}(r_1)$ under the condition that in both trial arms:

$$\text{corr}\left(y_{ij1}, p_{ij1}\right) > \frac{1 + \theta_i^2 - \left(\frac{\bar{c}_i}{R_i}\right)^2}{2\theta_i}. \tag{5}$$

Similarly, if r_1 and r_3 give the same estimate, $\text{Var}(r_3) < \text{Var}(r_1)$ under the condition that in both trial arms:

$$\text{corr}\left(y_{ij1}, p_{ij1}\right) > \theta_i/2. \tag{6}$$

If r_2 and r_3 give the same estimate, $\text{Var}(r_2) < \text{Var}(r_3)$ under the condition that in both trial arms $\text{Var}(R_i)/\bar{c}_i^2 < \text{Var}(R_i)/R_i^2$, that is,

$$\bar{c}_i/R_i > 1. \tag{7}$$

If $\text{corr}\left(y_{ij1}, p_{ij1}\right) = 0$, $\bar{c}_i/R_i > 1$ by Jensen's inequality. As we will see in the case study in Section 4, it is possible that $\text{corr}\left(y_{ij1}, p_{ij1}\right)$ approximately equals zero in real-world situations.

In contrast, with large n_1, n_0 and randomization, the double ratio of counts (r_4) provides an asymptotically unbiased estimator of the direct effect:

$$r_4 = \frac{R_1^*}{R_0^*} = \frac{\sum_{j=1}^{n_1} y_{1j1} / \sum_{j=1}^{n_1} y_{1j0}}{\sum_{j=1}^{n_0} y_{0j1} / \sum_{j=1}^{n_0} y_{0j0}}$$

$$\approx \frac{\sum_{j=1}^{n_1} \exp\left(\beta_D + \beta_I + \alpha_{1j} + \gamma\right) p_{1j1} / \sum_{j=1}^{n_1} \exp\left(\beta_I + \alpha_{1j}\right) p_{1j0}}{\sum_{j=1}^{n_0} \exp\left(\alpha_{0j} + \gamma\right) p_{0j1} / \sum_{j=1}^{n_0} \exp\left(\alpha_{0j}\right) p_{0j0}} = \exp\left(\beta_D\right).$$

The variance of r_4 is:

$$\text{Var}(r_4) = r_4^2 \left[\frac{\text{Var}\left(R_1^*\right)}{\left(R_1^*\right)^2} + \frac{\text{Var}\left(R_0^*\right)}{\left(R_0^*\right)^2} \right], \tag{8}$$

where

$$\frac{\text{Var}\left(R_i^*\right)}{\left(R_i^*\right)^2} = \frac{1}{n_i} \left[\text{CV}\left(y_{ij1}\right)^2 + \text{CV}\left(y_{ij0}\right)^2 - 2\text{CV}\left(y_{ij1}\right) \text{CV}\left(y_{ij0}\right) \text{corr}\left(y_{ij1}, y_{ij0}\right) \right].$$

Comparisons of the variances of the estimators r_2 vs r_4 and r_3 vs r_4 are meaningful only if the indirect effect is absent or trivial and the estimates $r_2 \cong r_4$ and $r_3 \cong r_4$. In Section A of Online Supplementary Material 1 we show that $\text{Var}(r_4) < \text{Var}(r_2)$ if in both trial arms:

$$\text{corr}\left(y_{ij1}, y_{ij0}\right) - \left(\frac{R_i}{\bar{c}_i}\right)^2 \theta_i \text{corr}\left(y_{ij1}, p_{ij1}\right) > 1 - \left(\frac{R_i}{\bar{c}_i}\right)^2 \times \frac{1 + \theta_i^2}{2}. \tag{9}$$

Similarly, $\text{Var}(r_4) < \text{Var}(r_3)$ if in both trial arms:

$$\text{corr}\left(y_{ij1}, y_{ij0}\right) - \theta_i \text{corr}\left(y_{ij1}, p_{ij1}\right) > \frac{1 - \theta_i^2}{2}. \tag{10}$$

A strong correlation between number of events in the target and non-target groups as compared to the correlation between the number of events and person-time in the target group would favor r_4 over r_2 and r_3 in terms of precision.

If a non-trivial indirect effect is present, the absolute values of the test statistics

$$|t(r_2)| = \frac{|r_2 - 1|}{\sqrt{\text{Var}(r_2)}},$$

$$|t(r_3)| = \frac{|r_3 - 1|}{\sqrt{\text{Var}(r_3)}}$$

and

$$|t(r_4)| = \frac{|r_4 - 1|}{\sqrt{\text{Var}(r_4)}}$$

are comparable in the sense that they all indicate the probability of rejecting the null hypothesis of the target ratio being one. Let $r_2 = \zeta r_4$, then $|t(r_2)| < |t(r_4)|$ if in both trial arms (details in Online Supplementary Material 1):

$$\text{corr}(y_{ij1}, y_{ij0}) > 1 - \zeta^2 \left(\frac{r_4 - 1}{\zeta r_4 - 1} \right)^2 \left(\frac{R_i}{\bar{c}_i} \right)^2 \left[\frac{1 + \theta_i^2}{2} - \theta_i \text{corr}(y_{ij1}, p_{ij1}) \right]. \tag{11}$$

Similarly, let $r_3 = \xi r_4$, then $|t(r_3)| < |t(r_4)|$ if in both trial arms:

$$\text{corr}(y_{ij1}, y_{ij0}) > 1 - \xi^2 \left(\frac{r_4 - 1}{\xi r_4 - 1} \right)^2 \left[\frac{1 + \theta_i^2}{2} - \theta_i \text{corr}(y_{ij1}, p_{ij1}) \right]. \tag{12}$$

In the special case that $\zeta = 1$ or $\xi = 1$, Equations (11) or (12) reduce to Equations (9) and (10), respectively. Otherwise, assume that the estimates of direct and indirect effects are in the same direction, the closer ζ or ξ is to 1, the more favourable r_4 is in terms of power.

Similar to r_4, the ratio of event rates estimator, $r_5 = R_1^\dagger / R_0^\dagger = \frac{R'_{11}/R'_{10}}{R'_{01}/R'_{00}}$, where $R'_{ik} = \frac{\sum_j y_{ijk}}{\sum_j p_{ijk}}$, also provides an asymptotically unbiased estimator of the direct effect. The variance of r_5 is:

$$\text{Var}(r_5) = r_5^2 \left[\frac{\text{Var}\left(R_1^\dagger\right)}{R_1^{\dagger 2}} + \frac{\text{Var}\left(R_0^\dagger\right)}{R_0^{\dagger 2}} \right] \tag{13}$$

where

$$\text{Var}\left(R_i^\dagger\right) = \text{Var}\left(\frac{R'_{i1}}{R'_{i0}} \right) = R_i^{\dagger 2} \left(\frac{\text{Var}(R'_{i1})}{R'^2_{i1}} + \frac{\text{Var}(R'_{i0})}{R'^2_{i0}} - \frac{2\text{cov}(R'_{i1}, R'_{i0})}{R'_{i1} R'_{i0}} \right),$$

$$\frac{\text{Var}(R'_{ik})}{R'^2_{ik}} = \frac{1}{n_i} \left[\text{CV}(y_{ijk})^2 + \text{CV}(p_{ijk})^2 - 2\text{CV}(y_{ijk})\text{CV}(p_{ijk})\text{corr}(y_{ijk}, p_{ijk}) \right],$$

$$\text{cov}(R'_{i1}, R'_{i0}) = \frac{1}{n_i \bar{p}_{i\cdot 1} \bar{p}_{i\cdot 0}} \left\{ \text{cov}(y_{ij1}, y_{ij0}) + R'_{i1} R'_{i0} \text{cov}(p_{ij1}, p_{ij0}) - R'_{i0} \text{cov}(y_{ij1}, p_{ij0}) - R'_{i1} \text{cov}(y_{ij0}, p_{ij1}) \right\},$$

and $\bar{p}_{i\cdot k} = \sum_{j=1}^{n_i} p_{ijk}/n_i$.

Furthermore,

$$\frac{\text{Var}\left(R_i^\dagger\right)}{R_i^{\dagger 2}} = \frac{\text{Var}(R'_{i1})}{R'^2_{i1}} + \frac{\text{Var}(R'_{i0})}{R'^2_{i0}} - \frac{2\text{cov}(R'_{i1}, R'_{i0})}{R'_{i1} R'_{i0}}$$

$$= \frac{1}{n_i} \left\{ \text{CV}(y_{ij1})^2 + \text{CV}(p_{ij1})^2 - 2\text{CV}(y_{ij1})\text{CV}(p_{ij1})\text{corr}(y_{ij1}, p_{ij1}) \right.$$

$$+ \text{CV}(y_{ij0})^2 + \text{CV}(p_{ij0})^2 - 2\text{CV}(y_{ij0})\text{CV}(p_{ij0})\text{corr}(y_{ij0}, p_{ij0})$$

$$- 2 \left[\text{CV}(y_{ij1})\text{CV}(y_{ij0})\text{corr}(y_{ij1}, y_{ij0}) + \text{CV}(p_{ij1})\text{CV}(p_{ij0})\text{corr}(p_{ij1}, p_{ij0}) \right.$$

$$\left. \left. - \text{CV}(y_{ij1})\text{CV}(p_{ij0})\text{corr}(y_{ij1}, p_{ij0}) - \text{CV}(y_{ij0})\text{CV}(p_{ij1})\text{corr}(y_{ij0}, p_{ij1}) \right] \right\}.$$

Then, $\text{Var}(r_5) < \text{Var}(r_4)$ if in both trial arms (details in Online Supplementary Material 1):

$$\text{corr}(y_{ij1}, p_{ij0}) + \text{corr}(y_{ij0}, p_{ij1}) < \text{corr}(y_{ij0}, p_{ij0}) + \text{corr}(y_{ij1}, p_{ij1}). \tag{14}$$

It is natural to expect that the number of events is more strongly correlated with the person-time in the same group than the other group. Therefore, we anticipate a high chance of $\text{Var}(r_5) < \text{Var}(r_4)$ in many studies.

For matched-pair CRTs, $n_1 = n_0 = n$ is the number of pairs of clusters. Within the jth pair of clusters, one cluster is randomized to receive intervention ($i = 1$) and the other is the control cluster ($i = 0$). The paired design version of the five estimators, r_l^{paired} ($l = 1, 2, 3, 4, 5$), and their variances are shown in Appendix Table A1.

2.3 | Approximately unbiased estimators

The literature about bias in ratio estimators and the mitigation methods has very much focused on paired observations, mostly concerning an estimator in the form of r_1^{paired}.[7,12] Rao and Pereira considered a ratio-of-ratio estimator in the form of r_3^{paired} or r_4^{paired}.[13] These previous works showed that the estimators have a bias of order n^{-1}; bias-reduction methods were proposed. Useful though they are, they do not deal with non-matched CRTs and r_2^{paired} and r_5^{paired}.

One solution is to determine the expectation and therefore bias of a ratio estimator, and then subtract the bias from the estimator. See, for example, van Kempen and van Vliet[8] and Rao and Pereira.[13] Although it has only been considered in studies of paired observations, the concept is applicable to both non-matched and matched-pair CRTs. Following this approach, we propose a set of approximately unbiased estimators. The key results for non-matched CRTs are shown below. Their matched-pair counterparts and details of the derivations are available in Section B of Online Supplementary Material 1.

2.3.1 | Ratio of mean counts in non-matched CRTs

The expectation of the asymptotically unbiased estimator r_1 and approximately unbiased estimator r_1^* of the ratio of means estimator are, respectively:

$$E(r_1) \approx \frac{\overline{Y}_{1\cdot1}}{\overline{Y}_{0\cdot1}}\left[1 + \frac{\text{Var}(y_{0j1})}{n_0 \overline{Y}_{0\cdot1}^2}\right],$$

$$r_1^* = r_1 - \left[E(r_1) - \frac{\overline{Y}_{1\cdot1}}{\overline{Y}_{0\cdot1}}\right] = r_1 - \frac{\overline{Y}_{1\cdot1}}{\overline{Y}_{0\cdot1}} \times \frac{\text{Var}(y_{0j1})}{n_0 \overline{Y}_{0\cdot1}^2} \approx r_1\left[1 - \frac{1}{n_0}\text{CV}^2(y_{0j1})\right],$$

with the unknown population mean $\overline{Y}_{0\cdot1}$ approximated by the sample mean $\overline{y}_{0\cdot1}$ to form the sample CV.

2.3.2 | Ratio of mean cluster-level event rates in non-matched CRTs

The expectation of the asymptotically unbiased estimator r_2 and approximately unbiased estimator r_2^* are, respectively:

$$E(r_2) \approx \frac{\overline{C}_1}{\overline{C}_0}\left[1 + \frac{\text{Var}(c_{0j})}{n_0 \overline{C}_0^2}\right], r_2^* = r_2 - \left[E(r_2) - \frac{\overline{C}_1}{\overline{C}_0}\right] \approx r_2\left[1 - \frac{1}{n_0}\text{CV}^2(c_{0j})\right], \tag{15}$$

with the unknown population mean of cluster event rates $\overline{C}_0$ replaced by the sample estimate to form the sample CV.

2.3.3 | Ratio of event rates in non-matched CRTs

$$E(r_3) \approx \frac{\overline{Y}_{1\cdot1}/\overline{P}_{1\cdot1}}{\overline{Y}_{0\cdot1}/\overline{P}_{0\cdot1}}\left[1 - \frac{\text{cov}(y_{1j1}, p_{1j1})}{n_1 \overline{Y}_{1\cdot1} \overline{P}_{1\cdot1}} - \frac{\text{cov}(y_{0j1}, p_{0j1})}{n_0 \overline{Y}_{0\cdot1} \overline{P}_{0\cdot1}} + \frac{\text{Var}(y_{0j1})}{n_0 \overline{Y}_{0\cdot1}^2} + \frac{\text{Var}(p_{1j1})}{n_1 \overline{P}_{1\cdot1}^2}\right],$$

$$r_3^* = r_3 - \left[E(r_3) - \frac{\overline{Y}_{1\cdot1}/\overline{P}_{1\cdot1}}{\overline{Y}_{0\cdot1}/\overline{P}_{0\cdot1}}\right]$$

$$\approx r_3\left[1 + \frac{1}{n_1}\text{CV}(y_{1j1})\,\text{CV}(p_{1j1})\,\text{corr}(y_{1j1}, p_{1j1}) + \frac{1}{n_0}\text{CV}(y_{0j1})\,\text{CV}(p_{0j1})\,\text{corr}(y_{0j1}, p_{0j1}) - \frac{1}{n_0}\text{CV}^2(y_{0j1}) - \frac{1}{n_1}\text{CV}^2(p_{1j1})\right],$$

with the unknown population mean $\overline{Y}_{i\cdot1}$ and $\overline{P}_{i\cdot1}$ ($i = 0, 1$) replaced by their sample estimates to form the sample CVs.

2.3.4 | Double ratio of counts in non-matched CRTs

$E(r_4)$ and r_4^* can be obtained by replacing r_3, p_{ij1} and $\overline{P}_{i\cdot1}$ by r_4, y_{ij0}, and $\overline{Y}_{i\cdot0}$ in the formula in the previous sub-section on ratio of event rates.

2.3.5 | Double ratio of event rates in non-matched CRTs

$$
E(r_5) \approx \frac{\frac{\overline{Y}_{1\cdot1}}{\overline{P}_{1\cdot1}} \Big/ \frac{\overline{Y}_{1\cdot0}}{\overline{P}_{1\cdot0}}}{\frac{\overline{Y}_{0\cdot1}}{\overline{P}_{0\cdot1}} \Big/ \frac{\overline{Y}_{0\cdot0}}{\overline{P}_{0\cdot0}}} \left\{ 1 - \sum_{i=0}^{1} \left[\frac{\text{cov}\left(y_{ij1}, y_{ij0}\right)}{n_i \overline{Y}_{i\cdot1} \overline{Y}_{i\cdot0}} + \frac{\text{cov}\left(p_{ij1}, p_{ij0}\right)}{n_i \overline{P}_{i\cdot1} \overline{P}_{i\cdot0}} + \sum_{k=0}^{1} \sum_{k'=0}^{1} \frac{(-1)^{k+k'}}{n_i} \frac{\text{cov}\left(y_{ijk}, p_{ijk'}\right)}{\overline{Y}_{i\cdot k} \overline{P}_{i\cdot k'}} \right] \right.
$$
$$
\left. + \frac{\text{Var}\left(y_{1j0}\right)}{n_1 \overline{Y}_{1\cdot0}^2} + \frac{\text{Var}\left(y_{0j1}\right)}{n_0 \overline{Y}_{0\cdot1}^2} + \frac{\text{Var}\left(p_{1j1}\right)}{n_1 \overline{P}_{1\cdot1}^2} + \frac{\text{Var}\left(p_{0j0}\right)}{n_0 \overline{P}_{0\cdot0}^2} \right\},
$$
$$
r_5^* \approx r_5 \left\{ 1 + \sum_{i=0}^{1} \left[\frac{1}{n_i} \left[\text{CV}\left(y_{ij1}\right) \text{CV}\left(y_{ij0}\right) \text{corr}\left(y_{ij1}, y_{ij0}\right) + \text{CV}\left(p_{ij1}\right) \text{CV}\left(p_{ij0}\right) \text{corr}\left(p_{ij1}, p_{ij0}\right) \right] \right. \right.
$$
$$
\left. \left. + \sum_{k=0}^{1} \sum_{k'=0}^{1} \frac{(-1)^{k+k'}}{n_i} \text{CV}\left(y_{ijk}\right) \text{CV}\left(p_{ijk'}\right) \text{corr}\left(y_{ijk}, p_{ijk'}\right) \right] - \frac{1}{n_0} \text{CV}^2\left(y_{0j1}\right) - \frac{1}{n_0} \text{CV}^2\left(p_{0j0}\right) - \frac{1}{n_1} \text{CV}^2\left(y_{1j0}\right) - \frac{1}{n_1} \text{CV}^2\left(p_{1j1}\right) \right\}.
$$

2.3.6 | Variances and confidence intervals

The variance of r_1^* is:

$$
\text{Var}\left(r_1^*\right) = E\left[r_1^* - E\left(r_1^*\right)\right]^2 = E\left\{ \left[r_1 - \frac{\overline{Y}_{1\cdot1}}{\overline{Y}_{0\cdot1}} \times \frac{\text{Var}\left(y_{0j1}\right)}{n_0 \overline{Y}_{0\cdot1}^2} \right] - \frac{\overline{Y}_{1\cdot1}}{\overline{Y}_{0\cdot1}} \right\}^2 = E[r_1 - E(r_1)]^2 = \text{Var}(r_1).
$$

Following the same steps, it can be shown that $\text{Var}\left(r_l^*\right) = \text{Var}(r_l)$ for $l = 2, 3, 4$, and 5 as well.

The distribution of ratios is not normal. For calculation of confidence intervals, we calculate $\ln\left(r_l^*\right) \forall l$. Using the delta method and the result above, $\text{Var}\left[\ln\left(r_l^*\right)\right] = \text{Var}(r_l)/r_l^{*2} \forall l$, where $\text{Var}(r_l)$ have been given in Equations (2)–(4), (8), and (13). Confidence intervals (CI) are calculated using the t-distribution with $n_1 + n_0 - 2$ degrees of freedom for non-matched CRTs and $n - 1$ degrees of freedom for matched-paired CRTs.[14] The CIs calculated are then exponentiated back to the original scale.

In Equation (4), the calculation of $\text{Var}(r_3)$ involves an asymptotic variance estimator of $\text{Var}(R_i) = \text{Var}\left(\sum_{j=1}^{n_i} y_{ij1}/\sum_{j=1}^{n_i} p_{ij1}\right)$. Similarly, in Equations (8) and (13), this variance estimator is involved in the calculation of $\text{Var}\left(R_i^*\right)$ and $\text{Var}\left(R_i^\dagger\right)$, and then the solutions are plugged into the estimators of $\text{Var}(r_4)$ and $\text{Var}(r_5)$, respectively. Cochran showed that this variance estimator gave a considerable under-estimation.[7] In contrast, he showed that the Jackknife method only mildly over-estimated the variance and the over-estimation vanished quickly as the number of observations increased. As such, an alternative method to statistical inference is to use the Jackknife method to estimate $\text{Var}(R_i)$, $\text{Var}\left(R_i^*\right)$, and $\text{Var}\left(R_i^\dagger\right)$ and then plug these values into the calculation of $\text{Var}\left(\ln\left(r_l^*\right)\right)$, $l = 3, 4$, and 5 and the respective CIs. We will use $r_{l(J)}^*$, $l = 3, 4$, and 5, to denote the estimators when used together with this Jackknife-based variance estimation method.

3 | SIMULATION

3.1 | Simulation setting

For non-matched CRTs, we generated the number of events in the kth group in the jth cluster in the ith trial arm, conditional on the person-time p_{ijk}, by using a Poisson distribution with expected value given by Equation (1). We considered three sets of intervention effects, representing (a) direct protection (with no indirect effect), (b) direct and indirect

protection, and (c) no effect, respectively: (a) $\exp(\beta_D) = 0.5$ and $\exp(\beta_I) = 1$; (b) $\exp(\beta_D) = 0.5$ and $\exp(\beta_I) = 0.75$; (c) $\exp(\beta_D) = 1$ and $\exp(\beta_I) = 1$. We set $\gamma = 1$ for the difference in event rate between the target and non-target groups. We set person-times (p_{ij1}, p_{ij0}) as following a bivariate distribution with means 100, coefficient of variation CV $(p_{ijk}) = 0.2$, 0.4, or 0.6, skewness 1.5, kurtosis 4, and correlation 0.8. Our choice of CV (p_{ijk}) takes into account the findings from a recent systematic review of CRTs that the first quartile, median, and third quartile of CV (p_{ijk}) were 0.22, 0.41, and 0.52, respectively.[15] We used positively skewed distributions because that is implied by the sizeable CV(0.6) and positive values of person-time. We used the *rmvnonnormal* macro in Stata for the non-normal data generation.[16] Additionally, we simulated person-time using a bivariate normal distribution with means 100 and CV $(p_{ijk}) = 0.2$ or 0.4, as symmetric distribution is possible under modest CV. A small number of observations (<1% in total) with person-time either below 5 or above 350 were replaced by 5 or 350, respectively. This is because CRTs often exclude clusters that are very small in size and exclude or sub-divide very large clusters due to operational and efficiency considerations.[5,6] We set the cluster effect, α_{ij}, as following a normal distribution with mean -2 and SD $(\alpha_{ij}) = 0.05$, 0.2, and 0.5. The three levels of SD (α_{ij}) correspond to approximately CV $(c_{ij}) = 0.05$, 0.2, and 0.5.[5] Hayes and Bennett noted that CV (c_{0j}) is "often ≤ 0.25 and seldom exceeds 0.5 for most health outcomes".[14]

For matched-pair CRTs, we generated the number of events in the ith trial arm in the jth paired cluster $(j = 1, 2, \ldots, n)$ by using a Poisson distribution with expected value $\exp(\beta_{D,ik} + \beta_{I,i} + \alpha_{ij} + \gamma) p_{ij1}$ in the target group and $\exp(\beta_{I,i} + \alpha_{ij}) p_{ij0}$ in the non-target group. We set the paired cluster effects $(\alpha_{1j}, \alpha_{0j})$ as following a bivariate normal distribution with means -2 and SDs 0.05, 0.2, or 0.5 and correlation 0.8. The person-time parameter in the kth group in the jth pair of clusters (p_{ijk}) followed a multivariate non-normal distribution with means 100, CV $(p_{ijk}) = 0.2$, 0.4, or 0.6, skewness 1.5, kurtosis 4, and correlation 0.8. Additionally, multivariate normal distribution was used in the case of CV $(p_{ijk}) = 0.2$ or 0.4. The other parameters in the matched-pair CRTs were the same as those in the non-matched CRTs.

In the literature, it has been suggested that non-matched CRTs should include at least four clusters per trial arm and matched-pair CRTs should include at least six pairs of clusters.[1] For non-matched CRTs, we evaluated the properties of the estimators when the number of clusters per trial arm is 4, 6, 8, 12, 16, 32, and 64. For matched-pair CRTs, we considered 6, 8, 12, 16, 32, and 64 pairs of clusters.

In each scenario, we conducted 10 000 replicates of data generation and in each of them calculated the five asymptotically unbiased estimators and the five approximately unbiased estimators and their variances. We report the relative bias of the mean estimates of the incidence rate ratio, root mean square error (RMSE), coverage probability (CP) of the 95% confidence intervals (CI) and power to reject the null hyperthesis of the respective ratio equals one (or type 1 error when the null hyperthesis is true). Calculation of CIs was based on log-transformation and then exponentiate back to the original scale, which were used for statistical inference. Calculations of relative bias and RMSE were based on the ratios themselves without transformation.

3.2 | Simulation results

In Figures 1 to 3 we show the simulation results of non-matched CRTs that the intervention only had a direct effect, that is, $\exp(\beta_D) = 0.5$ and $\exp(\beta_I) = 1$, and CV $(p_{ij}) = 0.4$, which was approximately the median level of variability in cluster size found by a systematic review.[15] To maintain visual clarity, we separately present the results on $r_1, r_2, r_3, r_1^*, r_2^*$, and r_3^* (upper panel) and r_4, r_5, r_4^*, and r_5^* (lower panel).

Figure 1 shows the patterns of relative bias. All the asymptotically unbiased estimators showed positive relative bias that decreased as the number of clusters per arm (n_i) increased. Furthermore, the bias of r_1, r_2, and r_3 increased as SD (α_{ij}) increased. In scenarios with SD $(\alpha_{ij}) \leq 0.2$, r_2, and r_3 only showed very mild bias and the two curves mostly overlapped. In contrast, the approximately unbiased estimators, r_1^*, r_2^*, and r_3^*, were practically unbiased under all situations considered. The relative bias of r_4 and r_5 was stable in relation to SD (α_{ij}). The approximately unbiased estimators r_4^* and r_5^* were practically unbiased under all situations considered.

In Figure 2, r_2^* and r_3^* had smaller RMSE than r_1^*, but they converged as n_i increased. r_2^* had slightly smaller RMSE than r_3^* when SD $(\alpha_{ij}) = 0.5$ and n_i was small. Otherwise they were almost indistinguishable from each other. r_5^* had smaller RMSE than r_4^*. Their difference was stable across level of SD (α_{ij}) but they converged as n_i increased. The asymptotically unbiased estimators had RMSE similar to or slightly larger than their respective approximately unbiased counterparts', but the difference vanished as n_i increased.

The estimators tended to have coverage probability (CP) smaller than the nominal 95% level, especially when SD (α_{ij}) was large (Figure 3). The CP improved as n_i increased. The asymptotically unbiased estimators tended to
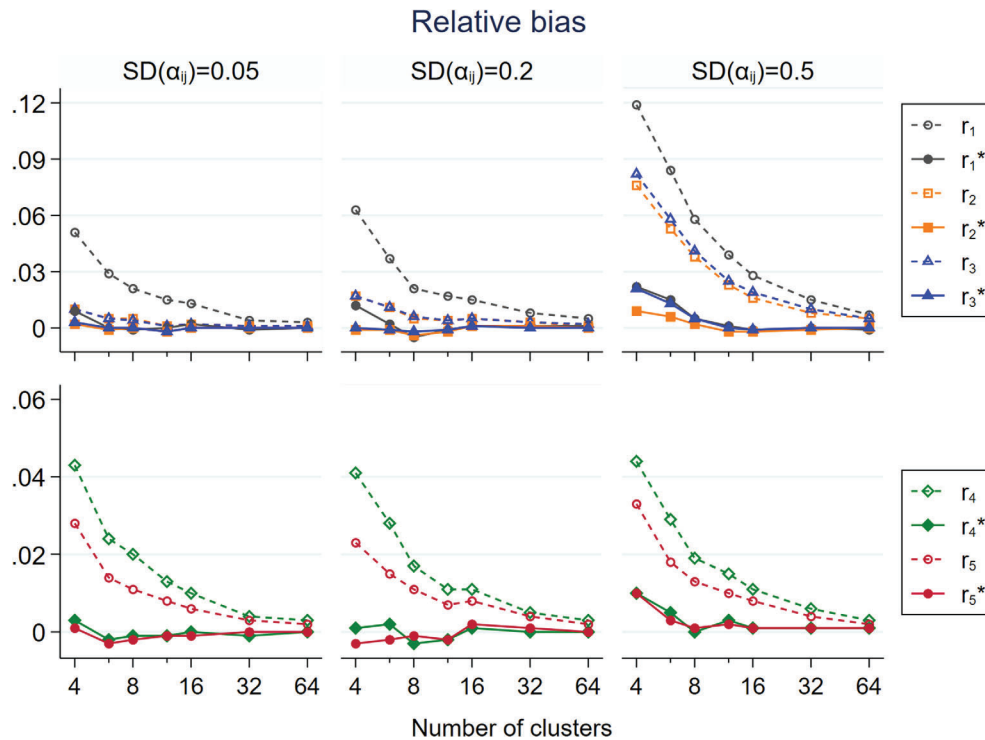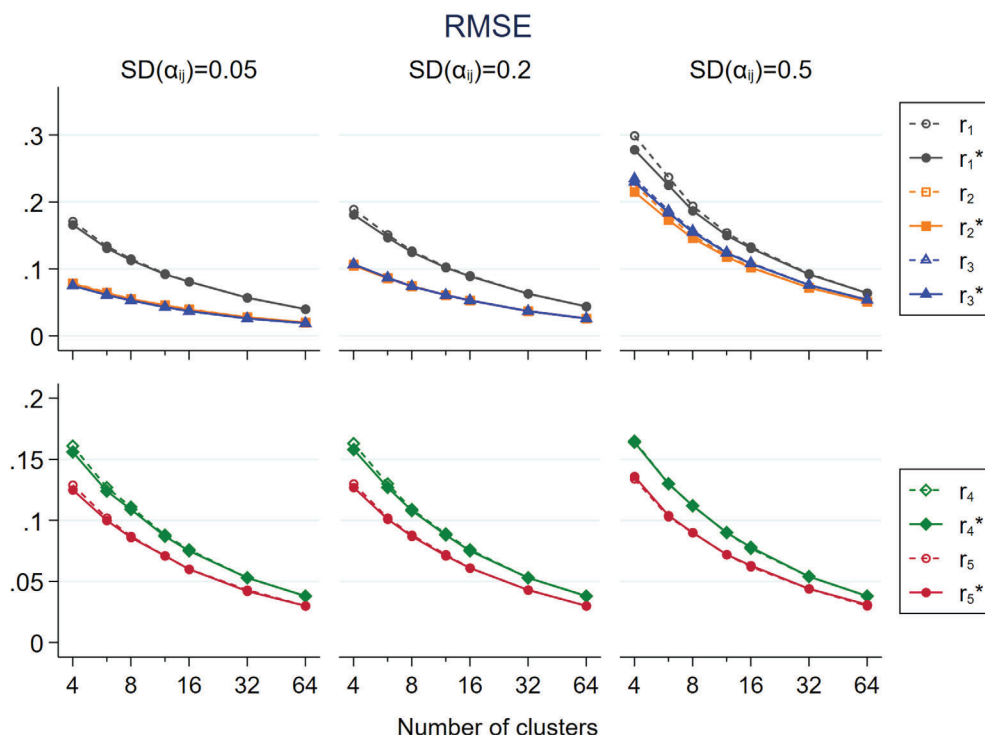
FIGURE 1 Relative bias of intervention effect estimators in relation to the number of clusters per trial arm for non-matched CRTs, by three levels of SD (α_{ij}); population size per cluster follows a skewed distribution with mean = 100 and CV = 0.4; intervention has a direct effect only $(\exp(\beta_D) = 0.5; \exp(\beta_I) = 1)$ [Colour figure can be viewed at wileyonlinelibrary.com]

have similar or slightly lower CP than their respective approximately unbiased counterparts. In the upper panel, r_2^* had CP close to the nominal 95% across all levels of n_i and SD (α_{ij}). r_1^* and r_3^* had CP below 94% in some scenarios of small n_i and large SD (α_{ij}). Using the Jackknife-based variance estimator described in Section 2 for r_3^*, denoted by $r_{3(J)}^*$, gave improved CP that was closer to the nominal 95% level than r_3^*. In the lower panel, r_4, r_5, r_4^*, and r_5^* had varying degree of under-coverage in different scenarios. However, using the Jackknife-based variance estimators for r_4^* and r_5^*, denoted by $r_{4(J)}^*$ and $r_{5(J)}^*$, respectively, the CP was close to the nominal level in all situations.

Figure 4 shows type 1 error rates, that is, rejection of null hypothesis in scenarios with $\exp(\beta_D) = 1$ and $\exp(\beta_I) = 1$. Otherwise, the parameters are the same as those in Figures 1 to 3. There was a tendency for all estimators to have type 1 error rate that exceeded the 5% target level, especially with SD $(\alpha_{ij}) \geq 0.2$ and $n_i < 32$. The inflation reduced as n_i increased. In the upper panel, r_2 and r_2^*, followed by $r_{3(J)}^*$ and r_1^*, performed better than the others. In the lower panel, r_4, r_5, r_4^*, and r_5^* had varying level of inflation of type 1 error under different parameter settings. In contrast, $r_{4(J)}^*$ and $r_{5(J)}^*$ performed well. In no circumstances did they show more than 1% deviation from the 5% target.

Figure 5 compares the power of selected estimators that use person-time as denominators, r_2^* and $r_{3(J)}^*$, vs selected estimators that use event counts in the non-target group as denominators, $r_{4(J)}^*$ and $r_{5(J)}^*$. We focused on them because they performed well in terms of CP and type 1 error rate. The lower panel introduced an indirect effect, $\exp(\beta_I) = 0.75$, in addition to a direct effect, $\exp(\beta_D) = 0.5$. Otherwise, the parameters here are the same as those in Figures 1 to 4. When there was direct effect only (upper panel), r_2^* and $r_{3(J)}^*$ were more powerful than $r_{4(J)}^*$ and $r_{5(J)}^*$ for SD $(\alpha_{ij}) \leq 0.2$. In contrast, $r_{4(J)}^*$ and $r_{5(J)}^*$ were more powerful when SD $(\alpha_{ij}) = 0.5$. Furthermore, $r_{5(J)}^*$ was more powerful than $r_{4(J)}^*$ in all the scenarios considered. With the addition of the indirect effect, r_2^* and $r_{3(J)}^*$ were more powerful than $r_{4(J)}^*$ at all levels of SD (α_{ij}), but similar to $r_{5(J)}^*$ at SD $(\alpha_{ij}) = 0.5$.

Further simulation results on non-matched CRTs under other parameter settings and simulation results on matched-pair CRTs are available in Online Supplementary Material 2. The findings are qualitatively similar to those reported above. Some relatively important additional information is as follows: First, in non-matched CRTs, the relative
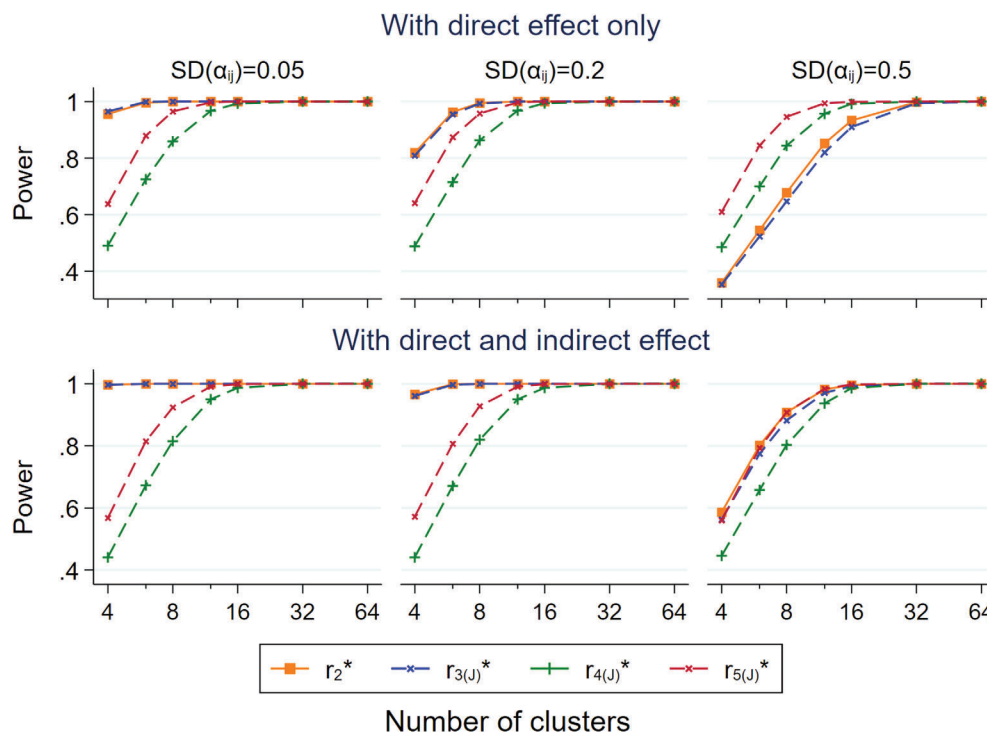
RMSE

FIGURE 2 Root mean squared error (RMSE) of intervention effect estimators in relation to the number of clusters per trial arm for non-matched CRTs, by three levels of SD (α_{ij}); population size per cluster follows a skewed distribution with mean = 100 and CV = 0.4; intervention has a direct effect only (exp $(\beta_D) = 0.5$; exp $(\beta_I) = 1$) [Colour figure can be viewed at wileyonlinelibrary.com]

CP
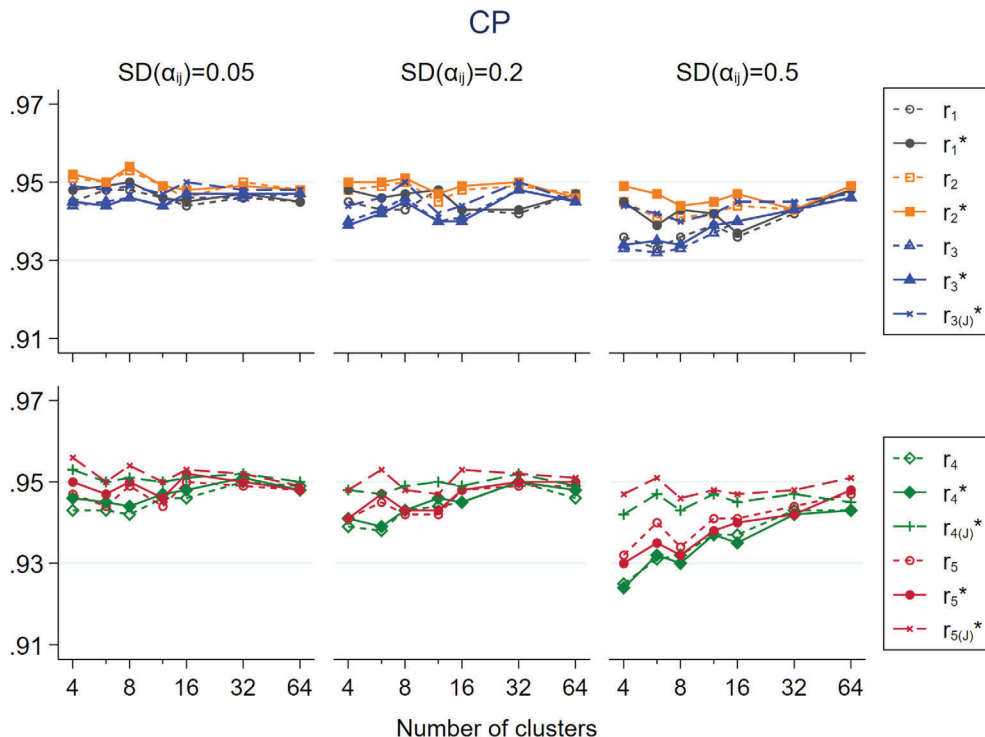
FIGURE 3 Coverage probability (CP) of 95% confidence interval (calculated on log-scale and exponentiated back to the original scale) in relation to the number of clusters per trial arm for non-matched CRTs, by three levels of SD (α_{ij}); population size per cluster follows a skewed distribution with mean = 100 and CV = 0.4; intervention has a direct effect only (exp $(\beta_D) = 0.5$; exp $(\beta_I) = 1$) [Colour figure can be viewed at wileyonlinelibrary.com]
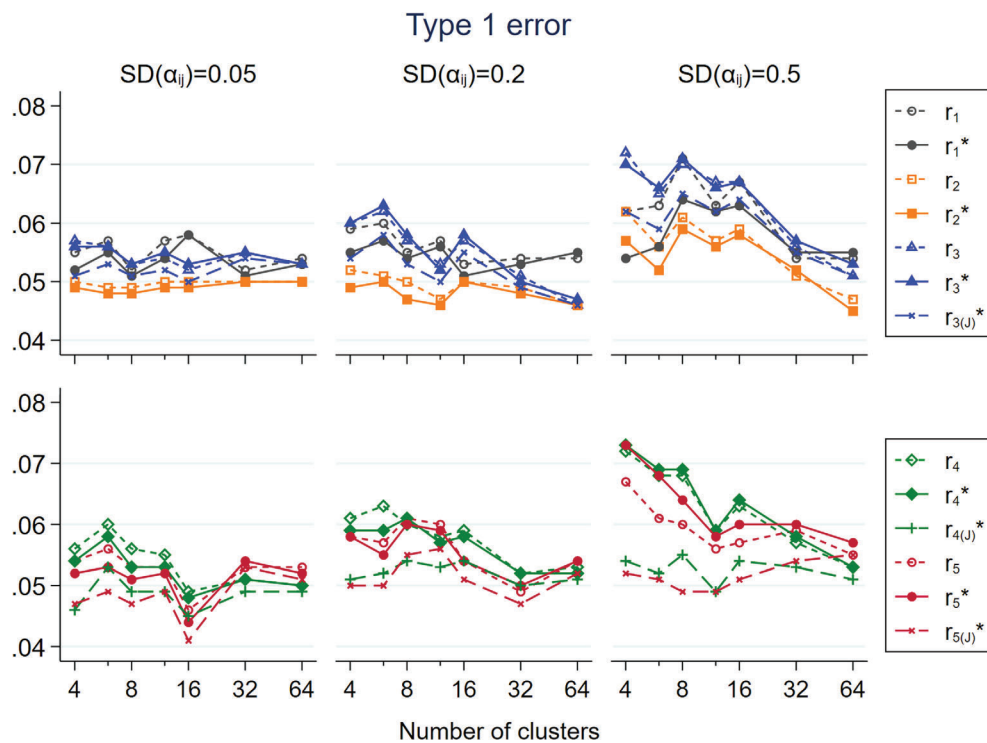
FIGURE 4   Type 1 error rate in relation to the number of clusters per trial arm for non-matched CRTs, by three levels of SD (α_{ij}); population size per cluster follows a skewed distribution with mean = 100 and CV = 0.4; intervention has no effect ($\exp(\beta_D) = 1; \exp(\beta_I) = 1$) [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 5   Power of $r_2^*, r_{3(J)}^*, r_{4(J)}^*$, and $r_{5(J)}^*$ in relation to the number of clusters per trial arm for non-matched CRTs, by three levels of SD (α_{ij}); population size per cluster follows a skewed distribution with mean = 100 and CV = 0.4. Upper panel: with direct effect only ($\exp(\beta_D) = 0.5; \exp(\beta_I) = 1$); lower panel: with direct and indirect effects ($\exp(\beta_D) = 0.5; \exp(\beta_I) = 0.75$) [Colour figure can be viewed at wileyonlinelibrary.com]

No. of clusters \ Years	2008	2009	2010
9			
18			
18			
9			

* light grey: children aged 3 to 59 months eligible for SMC;
dark grey: children aged 3 months to 9 years eligible for SMC;
white: no SMC.

FIGURE 6   Design of seasonal malaria chemoprevention (SMC) trial

bias of r_1 and r_4 increased substantially as CV (p_{ij1}) increased except when the number of cluster was about 32 or above. In contrast, r_2, r_3, r_5 and the approximately unbiased estimators were much less sensitive to the magnitude of CV (p_{ij1}) (eg, Figures S1 and S2). Second, in non-matched CRTs, the CP of $r_{3(J)}^*$ reduced to about 93% and its type 1 error rate increased to about 7% when CV $(p_{ij1}) = 0.6$, SD $(\alpha_{ij}) = 0.5$ and the number of clusters was 6 or below (eg, Figures S6 and S26). However, in matched-paired CRTs, $r_{3(J)}^{*,\text{paired}}$ performed well in these aspects while $r_2^{*,\text{paired}}$ had somewhat inflated type 1 error rate and below target CP (eg, Figures S41, S42, S62, and S63).

4 | SEASONAL MALARIA CHEMOPREVENTION TRIAL

We use a subset of data from a published study of seasonal malaria chemoprevention (SMC) in Senegalese children to illustrate.[17] The trial set-up is shown in Figure 6. The trial had a total of 54 clusters. It had the appearance of an "optimal design",[18,19] with nine clusters on intervention (leftmost column) and nine clusters on control condition (rightmost column) for all three time periods (2008 to 2010) that resembled a non-matched CRT, flanking a standard stepped-wedge trial (middle columns). The middle columns represent 18 clusters that were randomized to receive SMC from 2009 and another 18 clusters randomized to receive SMC from 2010. However, the trial was not planned according to the optimal design. The original plan was that the trial would continue up to 4 years (2008 to 2011), and the nine clusters on the rightmost column were randomized to receive SMC in 2011. But the trial was terminated after the malaria transmission season in 2010 according to data monitoring and interim analysis results. Furthermore, in the first period (2008), children aged 3 to 59 months in the nine clusters on the leftmost column were given SMC as part of the preparation of the study logistics. In 2009 and 2010, children aged between 3 months and 9 years (inclusive) were given SMC.

A passive surveillance system was implemented in health facilities to determine the number of clinical malaria episodes confirmed by rapid malaria test (primary endpoint) in each cluster, by four age groups (59 months or below; 5-9; 10-19; over 19 years). The data was at the cluster, not individual, level. Mortality data was collected for all age groups but only data for children aged 9 years or below was used in the previously published analysis[17] and available to the present analysis. Number of deaths (secondary endpoint) and population size of the clusters at mid-September each year (approximately the beginning of the annual malaria transmission season), by age groups, was collected by a demographic surveillance system.

For the purpose of illustration, we used the 2008 data from the 18 clusters that resembled a non-matched CRT. Children aged between 3 to 59 months were the target group. We considered children aged 5 to 9 years the non-target group.

Table 1 shows the descriptive statistics by endpoints and trial arms. It also included a simple average of each statistics in the two trial arms as a summary. There was very large between-cluster variability in malaria incidence rate in the target group, with CV $(c_{ij1}) > 1$ in both arms. In the target group, the CVs of malaria episodes approximately doubled the CVs of the population size, with $\theta = $ CV $(p_{ij1})/$CV $(y_{ij1}) \approx 0.60/1.37 = 0.44$. The correlation coefficient between malaria episodes and population size of the target group, corr (y_{ij1}, p_{ij1}), was weak, with average across trial arms being only 0.03. The estimates of malaria incidence rate in the SMC and control arms were $\bar{c}_1 = 0.0029 > R_1 = 0.0028$ and $\bar{c}_0 = 0.0031 > R_0 = 0.0020$, respectively. In contrast, the correlation between malaria episodes in the target and non-target group, corr (y_{ij1}, y_{ij0}), was strong, averaged at 0.76.

TABLE 1 Descriptive summary of number of malaria episodes and deaths, population size and their correlations and coefficient of variation of cluster-level event rates in 9 intervention clusters and 9 control clusters in a seasonal malaria chemoprevention (SMC) trial in 2008

Statistics	Malaria			Mortality		
	SMC	Control	Average	SMC	Control	Average
$\bar{y}_{ij1}$	8.00	3.56	5.78	7.11	2.89	5.00
$\bar{y}_{ij0}$	15.44	5.89	10.67	1.00	0.56	0.78
$\bar{p}_{ij1}$	2834	1741	2288	2834	1741	2288
$\bar{p}_{ij0}$	1977	1187	1582	1977	1187	1582
CV (c_{ij1})	1.21	1.75	1.48	0.89	0.68	0.79
CV (y_{ij1})	1.24	1.49	1.37	1.05	0.95	1.00
CV (y_{ij0})	1.15	1.08	1.12	1.58	1.31	1.45
CV (p_{ij1})	0.46	0.73	0.60	0.46	0.73	0.60
CV (p_{ij0})	0.49	0.65	0.57	0.49	0.65	0.57
corr (y_{ij1}, p_{ij1})	0.10	−0.05	0.03	0.48	0.69	0.59
corr (y_{ij1}, p_{ij0})	0.13	−0.05	0.04	0.57	0.68	0.63
corr (y_{ij0}, p_{ij1})	0.26	0.05	0.15	0.18	0.58	0.38
corr (y_{ij0}, p_{ij0})	0.34	0.01	0.17	0.26	0.57	0.41
corr (y_{ij1}, y_{ij0})	0.88	0.64	0.76	0.67	0.35	0.51
corr (p_{ij1}, p_{ij0})	0.988	0.998	0.993	0.988	0.998	0.993

Table 2 shows the estimation results. The estimates based on the asymptotically unbiased estimators were all larger than those based on the approximately unbiased estimators. Given this number of clusters, we considered the latter more accurate. For $l = 3$, 4, and 5, the SE's based on $r^*_{l(J)}$ were only slightly larger than r^*_l. All confidence intervals were quite wide. Although corr (y_{ij1}, p_{ij1}) was weak, the estimated r_1 was much larger than r_2 and r_3, leading to standard error SE (r_1) larger than SE (r_2) and SE (r_3). Given $\bar{c}_i > R_i$ in both trial arms, SE $(r_2) <$ SE (r_3). As expected from Equations (9) and (10), with strong corr (y_{ij1}, y_{ij0}) and weak corr (y_{ij1}, p_{ij1}), SE (r_4) was smaller than SE (r_2) and SE (r_3). The estimators for direct effects, r^*_4 and r^*_5, gave very similar result. Since there was no practical difference between corr $(y_{ij1}, p_{ij1}) +$ corr (y_{ij0}, p_{ij0}) and corr $(y_{ij1}, p_{ij0}) +$ corr (y_{ij0}, p_{ij1}) in either trial arm, as indicated by Equation (14), Var $(r_4) \cong$ Var (r_5).

Table 1 also shows that the descriptive statistics on mortality. There was less between-cluster variability in mortality rate than malaria incidence, with CV (c_{ij1}) in the two arms averaged at 0.79. Furthermore, in the target group, the CVs of death approximately doubled the CVs of the population size, with $\theta =$ CV (p_{ij1}) /CV $(y_{ij1}) \approx 0.60/1.00 = 0.60$. Unlike malaria episodes, the correlation coefficient between deaths and population size of the target group was more substantial, with average across trial arms being 0.59. The estimates of mortality rate in the SMC and control arms were $\bar{c}_1 = 0.0022 < R_1 = 0.0025$ and $\bar{c}_0 = 0.0018 > R_0 = 0.0017$, respectively. The correlation between deaths in the target and non-target group was moderate, with average corr (y_{ij1}, y_{ij0}) at 0.51; this was weaker than that for malaria episodes.

Estimation results on mortality are available in Table 2. Again, the asymptotically unbiased estimators tended to give larger estimates than their respective approximately unbiased estimators. However, the differences between the two set of estimates were smaller than those on malaria episodes, as expected from the smaller between-cluster variability in mortality rate than malaria incidence rate. The SE's based on $r^*_{l(J)}$ were only slightly larger than those based on r^*_l. All confidence intervals were quite wide. Since corr (y_{ij1}, p_{ij1}) was substantial and the estimated r_1 was much larger than r_2 and r_3, SE (r_2) and SE (r_3) were smaller than SE (r_1). With fairly similar values of corr (y_{ij1}, y_{ij0}) and corr (y_{ij1}, p_{ij1}) and relatively large θ, SE (r_2) and SE (r_3) were smaller than SE (r_4). There was no clear difference between corr $(y_{ij1}, p_{ij1}) +$ corr (y_{ij0}, p_{ij0}) and corr $(y_{ij1}, p_{ij0}) +$ corr (y_{ij0}, p_{ij1}) in either trial arm, so Var $(r_4) \cong$ Var (r_5).

TABLE 2 Estimates, SE and 95% confidence intervals (CI; exponentiation of log-transformed values) for malaria and mortality in seasonal malaria chemoprevention trial data in 2008

Endpoint	Estimator	Estimate	SE	95% CI
Malaria	r_1^*	1.70	1.45	(0.28, 10.4)
	r_2^*	0.62	0.67	(0.06, 6.07)
	r_3^*	1.01	0.97	(0.13, 7.80)
	$r_{3(J)}^*$	1.01	1.00	(0.12, 8.26)
	r_4^*	0.74	0.37	(0.25, 2.14)
	$r_{4(J)}^*$	0.74	0.40	(0.23, 2.32)
	r_5^*	0.76	0.38	(0.26, 2.19)
	$r_{5(J)}^*$	0.76	0.41	(0.24, 2.38)
	r_1	2.25	1.45	(0.57, 8.85)
	r_2	0.94	0.67	(0.21, 4.24)
	r_3	1.38	0.97	(0.31, 6.16)
	r_4	0.86	0.37	(0.34, 2.14)
	r_5	0.88	0.38	(0.35, 2.20)
Mortality	r_1^*	2.21	1.16	(0.73, 6.74)
	r_2^*	1.16	0.46	(0.50, 2.67)
	r_3^*	1.44	0.58	(0.61, 3.40)
	$r_{3(J)}^*$	1.44	0.61	(0.59, 3.54)
	r_4^*	1.08	0.81	(0.22, 5.26)
	$r_{4(J)}^*$	1.08	0.88	(0.19, 6.02)
	r_5^*	1.12	0.83	(0.23, 5.41)
	$r_{5(J)}^*$	1.12	0.91	(0.20, 6.33)
	r_1	2.46	1.16	(0.90, 6.70)
	r_2	1.22	0.46	(0.55, 2.70)
	r_3	1.51	0.58	(0.67, 3.42)
	r_4	1.37	0.81	(0.39, 4.78)
	r_5	1.40	0.83	(0.40, 4.92)

5 | DISCUSSION

The proposed approximately unbiased estimators successfully reduce the bias when the number of clusters is small. They also have advantages in terms of smaller RMSE and more accurate coverage probability than the asymptotically unbiased estimators. For studies with fewer than 60 clusters per arm, we recommend the use of the approximately unbiased estimators. Some CRTs do have a large number of clusters per trial arm, for example, a trial of influenza vaccination had over 400 nursing homes per arm[20] and a trial of mass drug administration had over 700 communities per arm.[21] For such studies, the choice between the asymptotically and approximately unbiased estimators is unimportant. Furthermore, with a large number of clusters, the simple estimator r_1 has performance very similar to r_2 and r_3. At the study planning stage, investigators may take into account this finding when they consider the cost and benefit of collecting person-time data.

Previous simulation studies evaluated the performance of the estimator r_2.[5,6] We caution that the range of parameter values they considered were somewhat narrow. From our analytic solution, the estimator is only asymptotically unbiased. As seen in Equation (15), the bias in the estimator r_2 is a non-linear function of the between-cluster variability of event

rate. From simulation, the bias of r_2 became obvious as the variability increased, especially when the number of clusters per trial arm was below 16 or so. In the SMC study of malaria episodes, where the variability was large, the r_2 estimate was much larger than r_2^*. In those situations, the use of our proposed bias-corrected estimator r_2^* is preferable over r_2.

While r_3^* performs well in terms of bias and RMSE, its variance estimator under-estimates the true variability as the CV of cluster size or event rate increases; it approaches the target CP as the number of clusters increases. The under-coverage can be corrected by plugging the Jackknife estimates of Var (R_i) into the estimator for Var (r_3). We found that r_2^* and $r_{3(J)}^*$ have similar performance except when CV $(p_{ijk}) = 0.6$ or SD $(\alpha_{ij}) = 0.5$. In those settings, $r_{3(J)}^*$ had type 1 error rate up to about 2% higher than the nominal 5% level when the number of clusters was six or below, and r_2^* may be preferred. However, in matched-pair CRTs, $r_2^{*,\text{paired}}$ did not out-perform $r_{3(J)}^{*,\text{paired}}$. Furthermore, previous studies had shown that the estimator of cluster-level event rate in a trial arm, $\bar{c}_i$, is a biased estimator, and the level of bias does not reduce in relation to increase in number of clusters.[8] In contrast, R_i can be used to obtain an asymptotically unbiased estimate of disease incidence.[7,8] Even if r_2^* is used to estimate incidence rate ratio, R_i is preferable over $\bar{c}_i$ as an estimator of the incidence rate in each trial arm.

Our consideration of r_4 and r_5 was in part motivated by an evaluation of a malaria vaccine where indirect effects were unlikely and the incidence of some outcomes such as meningitis (a safety outcome) could be highly variable between clusters, and where balanced randomization with respect to access to health care facilities (where passive surveillance took place) was difficult to ensure.[22] When SD (α_{ij}) is large, the number of events in the target group is likely to be more strongly correlated with the number of events in the non-target group than with the person-time in the target group. Furthermore, it is typical that vaccine studies anticipate no indirect effect on safety endpoints. So, these estimators are expected to estimate the same target as far as safety is concerned. In simulation studies of scenarios that the intervention only had a direct effect, we have seen that when SD (α_{ij}) was ≤ 0.2, r_2^* and $r_{3(J)}^*$ tended to be more powerful than $r_{4(J)}^*$ and $r_{5(J)}^*$. When SD (α_{ij}) increased to 0.5 and in the malaria data in the SMC trial, $r_{4(J)}^*$ and $r_{5(J)}^*$ had substantially smaller SE than r_2^* and $r_{3(J)}^*$. For situations like this, $r_{4(J)}^*$ and $r_{5(J)}^*$ are our estimators of choice. In various scenarios we evaluated, $r_{5(J)}^*$ gave higher level of statistical power than $r_{4(J)}^*$, except that they had similar performance when CV (p_{ijk}) was small. In that case, the magnitude of the benefits may not justify the extra cost in collection of person-time data in both the target and non-target groups. Otherwise, $r_{5(J)}^*$ tends to be preferable over $r_{4(J)}^*$.

We foresee that the estimators that use person-time in the target group and estimators that use number of events in the non-target groups as a denominator may be used in different parts of the same CRT, depending on the considerations aforementioned. For example, r_2 or r_3 or their extensions may be used in efficacy analysis while r_4 and r_5 or their extensions may be used in safety analysis.

Furthermore, r_4 and r_5 are generic quantities in the sense that y_{ij0} may be replaced by quantities other than the number of events in non-target groups to achieve other purposes. For example, there has been interest in the use of "negative control events" to remove the bias arising from differential ascertainment of outcome events in non-blinded CRTs.[4] The proposed estimator in the literature is in the form of r_4, with y_{1j0} and y_{0j0} replaced by the number of negative control events in the intervention and control arm, respectively.[4] The previous study did not consider the properties of the estimator in situations with small number of clusters. The results here apply directly. Another example of application is to the CRT with Before and after observations (CRT-BA) design,[23] which collects data in a baseline period before launching the randomized intervention and control comparator. By replacing y_{ij0} by the baseline event count, r_4 becomes a baseline adjusted estimator of the total effect. It offers a robust alternative to the analysis of CRT-BA trials.

The strengths of the present study include coverage of both non-matched and matched-pair CRTs, analytic evaluation of the bias of existing asymptotically unbiased estimators, proposal of estimators that capitalize on denominators other than cluster population size and their potential applications, development of a bias-corrected version of these estimators for use in studies with a small number of clusters, and simulation evaluation of the estimators with a realistic range of variability in cluster size. A limitation is that the methods do not handle covariate adjustment. However, good use of restricted randomization may reduce the need for covariate adjustment in the analysis stage.[1]

In non-matched CRTs, one approach for controlling covariate effects is stratified analysis and pooling of stratum-specific estimates using weights inversely proportional to variances. The limitation is that it is not practical to stratify for multiple covariates and it requires categorization of continuous covariates. Another approach is to apply ANCOVA to cluster-level data. However, it only works for methods that generate a summary value per cluster, such as r_1 and r_2, whose calculation begins with getting an event count and event rate per cluster, respectively. This is different from the calculation of r_3 to r_5, which begins with generating an estimate for a trial arm. For example, in r_3, the incidence rate in a trial arm is the sum of events over clusters divided by the sum of person-time over clusters within the trial arm.

There is not a summary value for every cluster. Furthermore, when the purpose is to estimate rate ratio instead of rate difference, the ANCOVA approach would need to analyze the log-transformed values instead. The exponentiated intervention effect estimate is then interpreted as a ratio of geometric means, which is not the same as the widely used estimator of ratio of arithmetic means (including r_1 and r_2). In CRTs with a small number of clusters, the bias and bias-correction for ratios of geometric means of event counts and event rates have yet to be investigated. Another approach is to use Poisson regression analysis with cluster-level covariates (and individual-level covariates if available) without using the intervention variable as predictors to obtain an expected number of events for each cluster.[5] Comparison of the deviations of the observed from the expected number of events between trial arms then offers a covariate adjusted estimate for the intervention effect. Following this idea, if y_{ij0} in r_4 or p_{ijk} in r_5 are replaced by this expected number of events, they become covariate adjusted estimators of the total or direct effect, respectively. The bias-correction method may then be applied to obtain covariate adjusted r_4^* or r_5^*. But the variance estimators may not work well as they do not account for the uncertainty in the prediction of the expected number of events. In short, while there are several candidate approaches available, challenges remain. Further research is needed to develop and evaluate these or other approaches to covariate adjustment.

The size of corr (y_{ij1}, p_{ij1}) is important in determining the relative strength of the different estimators. A large SD (α_{ij}) tends to dilute this correlation. As such, pilot data and careful consideration of the between-cluster variability in event rate is important not only for sample size determination but also for choice of study design and statistical analysis procedures.

CONFLICT OF INTEREST
The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

DATA AVAILABILITY STATEMENT
The data used in the case study are available at https://doi.org/10.17037/DATA.117. Requests for access will be reviewed by a Data Access Committee to ensure use of the data protect participant privacy according to the terms of participant consent and ethics committee approval. Stata codes for the simulations will be deposited to figshare by Wiley.

ORCID
Paul Milligan https://orcid.org/0000-0003-3430-3395
Kwok Fai Lam https://orcid.org/0000-0001-5453-994X
Yin Bun Cheung https://orcid.org/0000-0003-0517-7625

REFERENCES
1. Hayes RJ, Moulton LH. *Cluster Randomised Trials*. Boca Raton, FL: CRC Press; 2009.
2. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. Chichester, UK: Wiley; 2010.
3. Dron L, Taljaard M, Cheung YB, et al. Global health clinical trials: role and challenges of cluster randomised trials for global health. *Lancet Global Health*. 2021;9:e701-e710.
4. Dufault SM, Jewell NP. Analysis of counts for cluster randomized trials: negative controls and test-negative designs. *Stat Med*. 2020;39(10):1429-1439.
5. Bennett S, Parpia T, Hayes R, Cousens S. Methods for the analysis of incidence rates in cluster randomized trials. *Int J Epidemiol*. 2002;31(4):839-846.
6. Pacheco GD, Hattendorf J, Colford JM, Mausezahl D, Smith T. Performance of analytical methods for overdispersed counts in cluster randomized trials: sample size, degree of clustering and imbalance. *Stat Med*. 2009;28:2989-3011.
7. Cochran WG. *Sampling Techniques*. 3rd ed. Chichester, NY: Wiley; 1977.
8. van Kempen GMP, van Vliet LJ. Mean and variance of ratio estimators used in fluorescence ratio imaging. *Cytometry*. 2000;39:300-305.
9. Donner A, Klar N. Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *Am J Epidemiol*. 1994;140:279-289.
10. Rao JNK, Scott AJ. A simple method for the analysis of clustered binary data. *Biometrics*. 1992;48(2):577-585.
11. Halloran ME, Longini IM, Struchiner CJ. *Design and Analysis of Vaccine Studies*. New York, NY: Springer; 2010.
12. Durbin J. A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika*. 1959;46:477-480.
13. Rao JNK, Pereira NP. On double ratio estimators. *Sankhyā Ind J Stat Ser A*. 1968;30:83-90.
14. Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. *Int J Epidemiol*. 1999;28(2):319-326.
15. Kristunas C, Morris T, Gray L. Unequal cluster sizes in stepped-wedge cluster randomised trials: a systematic review. *BMJ Open*. 2017;7:e017151.
16. Lee S. Generating univariate and multivariate nonnormal data. *Stata J*. 2015;15(1):95-109.

17. Cisse B, Ba EH, Sokhan C, et al. Effectiveness of seasonal malaria chemoprevention in children under ten years of age in Senegal: a stepped-wedge cluster-randomized trial. *PLoS Med*. 2016;13(11):e1002175.

18. Girling AJ, Hemming K. Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. *Stat Med*. 2016;35:2149-2166.

19. Thompson JA, Fielding K, Hargreaves J, Copas A. The optimal design of stepped wedge trials with equal allocation to sequences and a comparison to other trial designs. *Clin Trials*. 2016;14(6):639-647.

20. Gravenstein S, Davidson HE, Taljaard M, et al. Comparative effectiveness of high-dose versus standard dose influenza vaccination on numbers of US nursing home residents admitted to hospital: a cluster-randomized trial. *Lancet Respir Med*. 2017;5(9):738-746.

21. Keenan JD, Bailey RL, West SK, et al. Azithromycin to reduce childhood mortality in sub-Saharan Africa. *New Engl J Med*. 2018;378(17):1583-1592.

22. Milligan P. *Statistical Analysis Plan for the Malaria Vaccine Pilot Evaluation*, London, UK: London School of Hygiene & Tropical Medicine and World Health Organization 2021. https://clinicaltrials.gov/ProvidedDocs/65/NCT03806465/SAP_001.pdf. Accessed July 11, 2021.

23. Hemming K, Taljaard M. Sample size calculations for stepped wedge and cluster randomised trials: a unified approach. *J Clin Epidemiol*. 2016;69:137-146.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

APPENDIX A.

TABLE A1 Estimators of incidence rate ratio for matched-pair cluster randomized trials[a]

Label	Estimand [b]	Estimator	Variance [c]
Ratio of means	$\exp(\beta_T)$	$r_1^{\text{paired}} = \dfrac{\bar{y}_1}{\bar{y}_0} = \dfrac{\sum_{j=1}^n y_{1j1}}{\sum_{j=1}^n y_{0j1}}$	$\dfrac{\bar{y}_1^2}{n\bar{y}_0^2}\left\{ \dfrac{\text{Var}(y_{1j1})}{\bar{y}_1^2} + \dfrac{\text{Var}(y_{0j1})}{\bar{y}_0^2} - \dfrac{2\text{cov}(y_{1j1},y_{0j1})}{\bar{y}_1\bar{y}_0} \right\}$
Ratio of mean cluster-level event rates	$\exp(\beta_T)$	$r_2^{\text{paired}} = \dfrac{\bar{c}_1}{\bar{c}_0} = \dfrac{\sum_{j=1}^n c_{1j}}{\sum_{j=1}^n c_{0j}}$	$\dfrac{\bar{c}_1^2}{n\bar{c}_0^2}\left\{ \dfrac{\text{Var}(c_{1j})}{\bar{c}_1^2} + \dfrac{\text{Var}(c_{0j})}{\bar{c}_0^2} - \dfrac{2\text{cov}(c_{1j},c_{0j})}{\bar{c}_1\bar{c}_0} \right\}$
Ratio of event rates	$\exp(\beta_T)$	$r_3^{\text{paired}} = \dfrac{R_1}{R_0} = \dfrac{\sum_{j=1}^n y_{1j1}/\sum_{j=1}^n p_{1j1}}{\sum_{j=1}^n y_{0j1}/\sum_{j=1}^n p_{0j1}}$	$\left(\dfrac{R_1}{R_0}\right)^2\left(\dfrac{\text{Var}(R_1)}{R_1^2} + \dfrac{\text{Var}(R_0)}{R_0^2} - \dfrac{2\text{cov}(R_1,R_0)}{R_1 R_0} \right)$
Double ratio of counts	$\exp(\beta_D)$	$r_4^{\text{paired}} = \dfrac{R_1^*}{R_0^*} = \dfrac{\sum_{j=1}^n y_{1j1}/\sum_{j=1}^n y_{1j0}}{\sum_{j=1}^n y_{0j1}/\sum_{j=1}^n y_{0j0}}$	$\left(\dfrac{R_1^*}{R_0^*}\right)^2\left(\dfrac{\text{Var}(R_1^*)}{R_1^{*2}} + \dfrac{\text{Var}(R_0^*)}{R_0^{*2}} - \dfrac{2\text{cov}(R_1^*,R_0^*)}{R_1^* R_0^*} \right)$
Double ratio of event rates	$\exp(\beta_D)$	$r_5^{\text{paired}} = \dfrac{R_1^{\dagger}}{R_0^{\dagger}} = \dfrac{\left(\frac{\sum_{j=1}^n y_{1j1}}{\sum_{j=1}^n p_{1j1}}\right)\Big/\left(\frac{\sum_{j=1}^n y_{1j0}}{\sum_{j=1}^n p_{1j0}}\right)}{\left(\frac{\sum_{j=1}^n y_{0j1}}{\sum_{j=1}^n p_{0j1}}\right)\Big/\left(\frac{\sum_{j=1}^n y_{0j0}}{\sum_{j=1}^n p_{0j0}}\right)}$	$\left(\dfrac{R_1^{\dagger}}{R_0^{\dagger}}\right)^2\left(\dfrac{\text{Var}(R_1^{\dagger})}{R_1^{\dagger 2}} + \dfrac{\text{Var}(R_0^{\dagger})}{R_0^{\dagger 2}} - \dfrac{2\text{cov}(R_1^{\dagger},R_0^{\dagger})}{R_1^{\dagger} R_0^{\dagger}} \right)$

[a] y_{ijk}, p_{ijk}, and $c_{ij} = y_{ij1}/p_{ij1}$ are, respectively, the number of events, person-time/population size and cluster-level event rates in the cluster that is randomized to receive the ith trial arm (1 for intervention and 0 for control) in the jth pair of clusters and kth group (1 for target and 0 for non-target group), and $\bar{c}_i = \sum_{j=1}^n c_{ij}/n$.

[b] β_T and β_D: Total effect and direct effect in terms of log incidence rate ratio.

[c] Components of the variances of r_l^{paired} $(l = 3, 4, 5)$ are: $\text{Var}(R_i) = \frac{1}{n\bar{p}_{i\cdot 1}^2}\left\{ \text{Var}(y_{ij1}) + R_i^2\text{Var}(p_{ij1}) - 2R_i\text{cov}(y_{ij1},p_{ij1}) \right\}, i = 0,1;$

$\text{cov}(R_1,R_0) = \frac{1}{n\bar{p}_{1\cdot 1}\bar{p}_{0\cdot 1}}\left\{ \text{cov}(y_{1j1},y_{0j1}) + R_1 R_0\text{cov}(p_{1j1},p_{0j1}) - R_1\text{cov}(y_{0j1},p_{1j1}) - R_0\text{cov}(y_{1j1},p_{0j1}) \right\},$

$\text{Var}(R_i^*) = \frac{1}{n\bar{y}_{i\cdot 0}^2}\left\{ \text{Var}(y_{ij1}) + R_i^{*2}\text{Var}(y_{ij0}) - 2R_i^*\text{cov}(y_{ij1},y_{ij0}) \right\}, i = 0,1;$

$\text{cov}(R_1^*,R_0^*) = \frac{1}{n\bar{y}_{1\cdot 0}\bar{y}_{0\cdot 0}}\left\{ \text{cov}(y_{1j1},y_{0j1}) + R_1^* R_0^*\text{cov}(y_{1j0},y_{0j0}) - R_1^*\text{cov}(y_{0j1},y_{1j0}) - R_0^*\text{cov}(y_{1j1},y_{0j0}) \right\},$

$\text{Var}(R_i^{\dagger}) = \text{Var}\left\{ \frac{R_{i1}'}{R_{i0}'} \right\} = \left(\frac{R_{i1}'}{R_{i0}'}\right)^2\left(\frac{\text{Var}(R_{i1}')}{R_{i1}'^2} + \frac{\text{Var}(R_{i0}')}{R_{i0}'^2} - \frac{2\text{cov}(R_{i1}',R_{i0}')}{R_{i1}' R_{i0}'} \right), i = 0,1;$

$\text{cov}(R_1^{\dagger},R_0^{\dagger}) = \text{cov}\left(\frac{R_{11}'}{R_{10}'}, \frac{R_{01}'}{R_{00}'}\right) = \frac{1}{nR_{10}'R_{00}'}\left\{ \text{cov}(R_{11}',R_{01}') + R_1^{\dagger}R_0^{\dagger}\text{cov}(R_{10}',R_{00}') - R_1^{\dagger}\text{cov}(R_{01}',R_{10}') - R_0^{\dagger}\text{cov}(R_{11}',R_{00}') \right\},$

$\text{Var}(R_{ik}') = \frac{1}{n\bar{p}_{i\cdot k}^2}\left\{ \text{Var}(y_{ijk}) + R_{ik}'^2\text{Var}(p_{ijk}) - 2R_{ik}'\text{cov}(y_{ijk},p_{ijk}) \right\}, \; i = 0,1; k = 0,1;$

$\text{cov}(R_{ik}',R_{i'k'}') = \frac{1}{n\bar{p}_{i\cdot k}\bar{p}_{i'\cdot k'}}\left\{ \text{cov}(y_{ijk},y_{i'jk'}) + R_{ik}'R_{i'k'}'\text{cov}(p_{ijk},p_{i'jk'}) - R_{ik}'\text{cov}(y_{i'jk'},p_{ijk}) - R_{i'k'}'\text{cov}(y_{ijk},p_{i'jk'}) \right\}, \; i,i' = 0,1; k,k' = 0,1;$ where

$R_{ik}' = \frac{\sum_{j=1}^n y_{ijk}}{\sum_{j=1}^n p_{ijk}}$ and $\bar{y}_{i\cdot k} = \sum_{j=1}^n y_{ijk}/n; \bar{p}_{i\cdot k} = \sum_{j=1}^n p_{ijk}/n, i = 0,1; k = 0,1.$