HHS Public Access

Author manuscript

*Stat Med*. Author manuscript; available in PMC 2020 August 26.

Published in final edited form as:

*Stat Med*. 2020 May 15; 39(10): 1429–1439. doi:10.1002/sim.8488.

# Analysis of counts for cluster randomized trials: Negative controls and test-negative designs

**Suzanne M. Dufault**[1], **Nicholas P. Jewell**[1,2]

[1]Division of Epidemiology and Biostatistics, University of California, Berkeley, California

[2]Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, United Kingdom

## Abstract

In cluster randomized trials (CRTs), the outcome of interest is often a count at the cluster level. This occurs, for example, in evaluating an intervention with the outcome being the number of infections of a disease such as HIV or dengue or the number of hospitalizations in the cluster. Standard practice analyzes these counts through cluster outcome rates using an appropriate denominator (eg, population size). However, such denominators are sometimes unknown, particularly when the counts depend on a passive community surveillance system. We consider direct comparison of the counts without knowledge of denominators, relying on randomization to balance denominators. We also focus on permutation tests to allow for small numbers of randomized clusters. However, such approaches are subject to bias when there is differential ascertainment of counts across arms, a situation that may occur in CRTs that cannot implement blinded interventions. We suggest the use of negative control counts as a method to remove, or reduce, this bias, discussing the key properties necessary for an effective negative control. A current example of such a design is the recent extension of test-negative designs to CRTs testing community-level interventions. Via simulation, we compare the performance of new and standard estimators based on CRTs with negative controls to approaches that only use the original counts. When there is no differential ascertainment by intervention arm, the count-only approaches perform comparably to those using debiasing negative controls. However, under even modest differential ascertainment, the count-only estimators are no longer reliable.

**Correspondence** Suzanne M. Dufault, Division of Epidemiology and Biostatistics, University of California, 2121 Berkeley Way West, Berkeley, CA 94720. sdufault@berkeley.edu.

AUTHOR CONTRIBUTIONS

N.P.J. proposed and advised the development of the contributed statistical method. S.M.D. helped develop the statistical method, performed the simulations, and wrote the first draft of the manuscript. Both authors reviewed and approved the final manuscript.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

DATA ACCESSIBILITY

The historical data used to inform the simulations in this article can be accessed in the supplementary material of our previous publication.[9]

SUPPORTING INFORMATION

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

**Keywords**

cluster randomization; dengue; health-care-seeking behavior; negative controls; permutation tests; test-negative design

## 1 | INTRODUCTION

Randomized controlled trials are the gold standard for evaluating the efficacy of health interventions. Randomization makes comparison groups as similar as possible in all factors except for the intervention under study and provides a basis of nonmodel based inference. When an intervention is delivered to groups of individuals, for example, in neighborhoods, or may have a community-wide health impact, randomization of the intervention necessarily occurs at the group, rather than individual, level. Such a trial is termed a cluster randomized trial (CRT).[1] The nonindependence of individuals within each cluster in CRTs causes statistical inefficiency—the "design effect"—necessitating inflation of the sample size to achieve power equivalent to an individually randomized trial.[1-3]

In many CRTs, outcome measurements are made at the cluster—rather than individual—level for a variety of reasons. For example, counts of events across a cluster may be collected by existing or designed surveillance systems. For CRT count outcome data, common estimators of the intervention effect include estimation of absolute and relative rate differences, usually based on demographic information on relevant population years of observation, or population size, per cluster.[1] When adjustment for cluster-level covariates is desirable, model-based regression modeling approaches are often used, including marginal generalized estimating equation (GEE) approaches and mixed effects models with random effects at the level of the cluster. The latter models can be extended to allow for individual-level information.

In many situations, these standard approaches require modification. For example, in settings where few clusters are available for randomization, model-based estimation and inference may be less accurate and require small sample size adjustments.[1] In such cases, a randomization-based strategy (eg, permutation tests) presents an attractive alternative. Furthermore, population-based denominators may sometimes be unavailable or not appropriate. The latter can occur when the count ascertainment system does not cover the entire cluster populations, perhaps due to access to care issues. The statistical analysis then depends solely on randomization balancing the unobserved population denominators across intervention arms. This risks unobserved bias—particularly in unblinded studies—due to differential ascertainment coverage across arms that will confound any intervention effect.

As noted, it is possible to estimate and test an intervention effect using only cluster-level case counts given intervention randomization. Here, we discuss such inference, focusing on the relative risk and its permutation distribution (under permuted intervention assignments). We subsequently consider the impact of differential ascertainment bias and introduce a method to remove, or reduce, such bias through the use of negative controls.[4] We discuss briefly the required properties for a valid negative-control count. We use simulations to address bias and precision comparisons between the various methods.

An example of a design that explicitly uses negative controls is the test-negative design (TND) that was recently extended to allow for cluster randomization of an intervention.[5] Test-negative designs are explicitly used to address ascertainment bias caused by differential health-care seeking behavior.[6,7] We thus interpret our findings in the context of cluster randomized TNDs with analytic methods that either only use case count data or use negative control (in addition to case count) information. Test-negative designs also directly accommodate the absence of population-level denominator information underlying the observed counts of interest.

These issues are motivated by the World Mosquito Program's ongoing balanced parallel-arm Cluster Randomized Test-Negative Design (CR-TND) trial to evaluate the efficacy of *Wolbachia*-infected mosquitoes in reducing the burden of dengue transmission in Yogyakarta City, Indonesia. In this study, Yogyakarta City, and its population of approximately 400,000, was divided into 24 contiguous clusters each measuring approximately 1 $km^2$ in size but with varying population density and socioeconomic status. Twelve of the clusters were randomly assigned to an intervention arm that received releases of *Wolbachia*-infected mosquitoes. *Wolbachia* successfully transinfected in nonnative hosts such as *Aedes aegypti* mosquitoes, the primary vectors of dengue, have been shown to disrupt the transmission of dengue and other flaviviruses by minimizing virus replication within the vector.[8] The remaining 12 clusters were assigned as control clusters. Count ascertainment depends on individuals seeking care at *puskesmas* (community health clinics) who present with general symptoms consistent with the clinical case definition of dengue. Such individuals who consent to enroll in the trial are subjected to laboratory testing for dengue, which determines their test-positive (case) or test-negative (control) status. The trial has been described in greater detail elsewhere.[9,10]

## 2 | DIRECT COMPARISON OF COUNTS IN THE ABSENCE OF POPULATION DENOMINATORS

We consider here a CRT for which the outcome is measured at the cluster level and comprises of a count of a number of “events" in each cluster. For example, the counts could represent the number of incident dengue infections over the study period as obtained through some well-defined ascertainment system. We let $A_j$ denote the observed count in the $j$th cluster assigned to the intervention and, analogously, $G_j$ is the count in the $j$th control cluster. Then, $A_T$ and $G_T$ are the total sum of the $j$ cluster-level case counts ($A_j$, $G_j$) in the treatment and control arms, respectively. That is, $A_T = \sum_{j=1}^{m} A_j$ and $G_T = \sum_{j=1}^{m} G_j$, where we assume, for convenience, that $m$ clusters are randomly assigned to both the intervention and control arms.

Given randomization, differences in the cluster counts between the intervention and the control arms should only arise through the intervention so long as case ascertainment is not differentially applied across arms. In particular, the underlying population denominators for a rate should be balanced across arms. Thus, to test the null hypothesis that there is no difference in the rate of case counts between the intervention and control arms, we can use the test statistic in Equation (1).

$$T = \sum A_j - \sum G_j = A_T - G_T \quad (1)$$

With small numbers of clusters, we focus on the permutation distribution of $T$ (across all permutations of clusters' intervention assignments). In the simple case considered here, there are $\binom{2m}{m}$ possible intervention assignments and computation of an estimate for each of these (while holding each cluster count fixed) yields the permutation distribution that can form the basis of randomization inference. It is immediate that $E_P[A_T] = E_P[G_T] = n_D/2$, where $E_P$ refers to the expectation under the permutation distribution and $n_D$ is the total of all counts across all clusters, that is, $A_T + G_T$, held fixed over all permutations. Furthermore, from finite sampling methods, $\mathrm{Var}_P(A_T) = mV_D/2$ where $V_D$ is the variance of the combined counts $A_1, \ldots, A_m, G_1, \ldots, G_m$ in the intervention and control clusters combined, with this variance calculated using $(2m - 1)$ in the denominator. This follows since, for a random permutation, the $A_j$ counts are simply randomly selected from the combined counts across all clusters.

Thus, $E_P[T] = E_P[A_T] - E_P[G_T] = 0$, and the permutation variance of $T$ is just $\mathrm{Var}_P(T) = 2mV_D$. Thus, to evaluate the null hypothesis of no intervention effect we can either use the full permutation distribution or approximate such an approach by comparing a standardized statistic, $T / \sqrt{2m\hat{V}_D}$—using an appropriate estimate of $V_D$—to a $t$ distribution with the appropriate number of degrees of freedom.

$V_D$ can be simply estimated by the empirical variance of the $A_j$s in the intervention clusters or the $G_j$s in the control clusters (or the variance of the counts combined across both arms). Since the arms contain the same number of clusters, a simple average of these two arm-specific variance estimates could be used, leading to the so-called pooled variance estimator for the two-sample $t$ test with $2(m - 1)$ as the appropriate number of degrees of freedom. The combined variance, and, to a lesser extent, the pooled estimator are likely to be biased in estimating $V_D$ in the presence of an intervention effect. This suggests an alternative approach when using the permutation distribution, or its approximation, as the basis for confidence intervals, which we discuss below.

We now turn to estimation of $\lambda$, the relative risk comparing intervention and control arms. One can think of $\lambda$ as the ratio of the underlying rates that generates the cluster counts in each arm. Alternatively, $\lambda$ is simply the ratio of the mean of the cluster counts across the two arms. Here, we focus on the estimator $\lambda_R = A_T/G_T = A_T/(n_D - A_T)$, where R simply stands for ratio (of the counts). For confidence intervals, we move to the symmetrically distributed version, $\log(\lambda_R)$. By definition, $E_P \log(\lambda_R) = 0$ at the null. Away from the null, we need to evaluate the permutation distribution of the $\log(\lambda_R)$ assuming an intervention effect. Note that the delta method can be used to approximate the permutation variance of $\log(\lambda_R) \approx (16 / n_D^2)(m / 2)V_D$.

Note that the intervention only affects the counts $A_1, \ldots, A_m$ by assumption. These are each replaced in turn by $A_1^*, \ldots, A_m^*$ which reflect altered counts in the intervention clusters. For

large populations, $A_j^* \approx \lambda A_j \approx \lambda A_j$ for the intervention clusters, assuming that the intervention effect is the same for all clusters. The common modification of the $A_1, \ldots, A_m$ has two immediate implications: first, under the permutation distribution, $E_P \log(\lambda_R)$ « $\log(\lambda)$ and second, there is no change to the variance formula $\log(\lambda_R) \approx (16 / n_D^2)(m / 2)V_D$ since all count ratios for different permutations are shifted by approximately $\log(\lambda)$.

However, away from the null, we have to modify the estimates of $n_D$, $V_D$ due to the replacement of each $A_j$ with $A_j^*$. The necessary adjustment is achieved by simply increasing the observed $A_j^*$s by the common factor $1/\lambda_R$ to obtain an estimate of $A_j$ (in the $j$ intervention clusters), en route to an estimate of $n_D$, $V_D$ as at the null.

## 3 | DIFFERENTIAL CASE ASCERTAINMENT

A fundamental threat to the validity of the approach of Section 2—even with randomization—arises when there is differential "counting" methods across the two arms. In such cases, when passive surveillance approaches are used to generate the necessary counts, differential case ascertainment may occur across treatment arms. For example, individuals' health-care-seeking behavior may be differential based on knowledge of their intervention assignment and this will affect any ascertainment system that is based on attendance in some health-care setting. This behavior is particularly relevant in trials where blinding of the participants and/or investigators to the intervention is infeasible for logistical, ethical, or other reasons. We refer to this phenomenon as differential count ascertainment. We stress that this threat to validity persists even if the relevant denominator information is known for the cluster counts.

We quantify this effect through the relative propensity $\pi$ of treated and untreated populations to "be counted", for example, seek health care. We allow this propensity to differ across treatment arms denoted by $E$ here, for convenience. That is, $E$ refers to individuals in the intervention arm and $\bar{E}$ to those in the control arm. Then we let

$$\alpha_{RA} = \frac{\Pr(A = 1 \mid E, D)}{\Pr(A = 1 \mid \bar{E}, D)},$$

where $A$ stands for ascertainment, RA for relative ascertainment, and the binary indicator $D$ denotes a "case" that would be counted if ascertainment was guaranteed.

It is obvious that with the comparison of counts across arms as described in Section 2, the effects of risk reduction and relative ascertainment are completely confounded and could not be disentangled without direct knowledge of $\alpha_{RA}$. One approach to address this fundamental bias is through use of negative controls. Negative controls, commonly used to calibrate measurements in laboratory experiments, have recently been reexamined for epidemiological applications.[4] The key requirements for a useful negative control outcome is that (a) no intervention effect is expected on the negative control outcome and (b) negative control outcomes must be affected by identical relative ascertainment effects as our outcome of interest. Note that the latter assumption allows differential ascertainment across intervention arms but this must occur in identical fashion as to what occurs for the outcome

of interest as quantified by $a_{\mathrm{RA}}$. It is exactly this assumption that allows estimation of $a_{\mathrm{RA}}$ and subsequent removal of ascertainment bias in estimation of $\lambda$.

The second of these conditions may appear difficult to achieve in any practical intervention study. We nevertheless introduce exactly such an example in the context of what are referred to as TNDs.

### 3.1 | The test-negative design

In infectious disease research, issues relating to differential case ascertainment, typically under the influence of differential health-care-seeking behavior, have been mitigated by the implementation of the TND. TNDs represent a variant on a traditional case-cohort design: studies enroll subjects who seek care for a clinical syndrome, defining those who test positive and negative for a pathogen of interest as "cases" and "controls," respectively. Specifically, the popularity of the TND arose from its ability to use existing surveillance systems (eg, clinic data) to estimate seasonal influenza vaccine effectiveness while minimizing bias due to health-care-seeking behavior. A nuanced discussion of this design can be found in the recent literature[6,11,12] that includes a formal analysis of causal diagrams associated with the design. The design and analytical methods were recently extended to cluster randomized interventions,[9,10] yielding the so-called cluster randomized test-negative design (CR-TND). A recent review of TNDs to mosquito vaccine effectiveness discusses 348 such studies.[13]

In a TND, test-positives play the role of our case counts in Section 2 and are ascertained through attendance, diagnosis, and testing at a clinic or other health-care setting. Subsequently, a critical component of the TND is the definition of test-negatives. As negative controls, the objective is to identify a disease that is unaffected by the intervention of interest and symptomatically similar to the disease outcome of interest. Upon recruitment at a clinic, a highly sensitive and specific laboratory test is used to distinguish test-positive cases (those with the disease of interest) from the test-negative controls (those without). The full extent of these assumptions have been critically discussed in the literature.[5,6] The key property of negative controls regarding differential ascertainment is explicitly achieved since participants do not know their disease status until they are ascertained and so it theoretically not possible for the test-positives and test-negatives to suffer from differential relative ascertainment, that is, the relative ascertainment $a_{\mathrm{RA}}$ is the same for $D$s (test-positives) as for $\bar{D}$s (test-negatives).

Using the cumulative notation provided in Table 1, that describes totals across clusters, the negative control assumption that the intervention has no impact on test-negatives leads to the proportion of test-negative individuals among the intervention care-seeking population ($B_T/N_{IO}$) being approximately equivalent to the proportion of test-negative individuals among the negative control care-seeking population ($H_T/N_{CO}$). Note that, in this context, $N_{IO}$ and $N_{CO}$ represent the unobserved denominators discussed in Section 2. It is then possible to approximate the natural, but unobserved, estimate of the relative risk of disease across the intervention and control populations $((A_T/N_{IO})/(G_T/N_{CO}))$ by substituting the ratio of test-negative individuals from the intervention and control subpopulations ($H_T/B_T$) as a proxy for

the unobserved relative sizes of the care-seeking intervention and control denominators ($N_{IO}/N_{CO}$). This results in the simple TND estimator, $\lambda_{\text{TND}} = A_T H_T / B_T G_T$.[6]

### 3.2 | Estimation of differential case ascertainment

Note that the assumptions of an appropriate negative control allow for estimation of the common relative ascertainment parameter $\alpha_{\text{RA}}$. The first assumption indicates that the relative counts of test-negatives in the intervention and control arms is not affected by the intervention which has a null effect on the negative control outcome. The second assumption then yields that the relative ascertainment of test-negatives is the same as for test-positives representing the outcome of interest.

Consider the scenario in which individuals within the intervention arm are ascertained differentially from individuals within the control arm. Focusing on the test-negatives, our assumptions show that $\alpha_{\text{RA}} = \Pr(A = 1 \mid E, \bar{D}) / \Pr(A = 1 \mid \bar{E}, \bar{D})$. Provided the other CR-TND assumptions hold,[5] and with the assumption of no intervention effect on the negative controls, $\alpha_{\text{RA}}$ can be estimated by the identical approach previously outlined for case count-only estimation of the RR, that is, $\hat{\alpha}_{\text{RA}} = B_T / H_T$. This provides an unbiased estimator of the relative ascertainment parameter.

The variance of $\hat{\alpha}_{\text{RA}}$ can be estimated exactly as we described for the intervention effect estimate in Section 2: $\text{Var}_p(\hat{\alpha}_{\text{RA}}) \approx (16 / n_{\bar{D}}^2)(m / 2) V_{\bar{D}}$, where $V_{\bar{D}}$ is the variance of the clusters' test-negative counts combined across intervention arms and $n_{\bar{D}} = B_T + H_T$. To assess whether ascertainment (of the negative controls) differs across arms, a suitable test statistic is, again, the difference in counts, $T = B_T - H_T$, scaled by the variance $\text{Var}_p(T) = 2m V_{\bar{D}}$, where $V_{\bar{D}}$ is the population variance of the $2m$ test-negative counts, and compared with a $t$ distribution with $2(m - 1)$ degrees of freedom (assuming we use a variance estimate that averages variability across the two arms as described in Section 2). This test is of interest in its own right when negative control information is available as it assesses differential ascertainment effects across arms independently of any intervention. Such information may be useful in planning and interpreting future trials.

### 3.3 | Estimating the intervention effect, $\lambda$, in the presence of differential case ascertainment

When $\alpha_{\text{RA}} \neq 1$, the estimated intervention effect given by $\lambda_{\text{R}}$ is necessarily biased, as noted above, that is, the estimate is shifted multiplicatively by $\alpha_{\text{RA}}$ (or, additively, by log $\alpha_{\text{RA}}$ on the log scale). Without further information, this reflects the vulnerability to bias of the "count-only" approach of Section 2. However, knowledge of the negative control counts allows estimation of $\alpha_{\text{RA}}$ as shown in Section 3.2. Thus, a "debiased" intervention relative risk can then be estimated by $\hat{\lambda} = \lambda_{\text{R}} \times \alpha_{\text{RA}}^{-1} = \frac{A_T H_T}{G_T B_T}$. This, of course, is precisely the simple TND estimator ($\lambda_{\text{TND}}$) proposed for all TNDs including the CR-TND. Randomization-based inference associated with this estimator is presented in previous work.[9]

## 4 | SIMULATIONS

Data-based simulations evaluate the performance of the proposed estimation methods. As a practical basis for simulations, historical counts of dengue from 24 contiguous clusters within a city in Indonesia collected from 2003 to 2014 were divided into nine consecutive* 2-year periods. Other febrile illnesses (OFIs) with similar presenting symptoms will be used as negative controls. Counts of OFIs for each of the 24 clusters from 2014 to 2015 provided the historical distribution of these negative controls. Exact distributions of these historical counts can be found in supplemental material previously published.[9] For each historical period, complete random assignment was performed such that $m = 12$ of the total 24 clusters were assigned to a putative intervention and the remainder to control.

Instead of building an exhaustive permutation distribution of the more than two million distinct intervention allocations for each time period, each simulation assigned intervention according to the same 10,000 distinct potential intervention allocations and examined the results of these intervention allocations across all nine historical time periods.

For a specific period, the distribution of the case counts ($n_D$) and negative control counts ($n_{\bar{D}}$) among clusters is assumed to follow multinomial distributions parameterized by the observed historical cluster-level proportions of cases (or negative controls) that fell in cluster $j$, $p_{Dj}$, or $p_{\bar{D}j}$, respectively. Given an intervention effect $\lambda$, $p^*_{Dj} = \lambda p_{Dj}$ for all clusters in the intervention arm with the other proportions in the control cluster left unchanged. These adjusted proportions are then standardized such that $\sum_{i=1}^{2m} \mathrm{I}(E = 1) \times \lambda p_{Di} + \{1 - \mathrm{I}(E = 1)\} \times p_{Di} = 1$. The negative control distribution is unaffected by the intervention by definition.

To allow for potential differential ascertainment by intervention arm, we assume that $\alpha_{\mathrm{RA}}$ can be applied in a similar manner except that it also modifies the distribution of negative controls. Since $\alpha_{\mathrm{RA}}$ is a relative measure of differential ascertainment, we modify all case counts and negative control counts within the intervention arm only. After this modification, the proportions are again standardized such that the proportions of case counts and negative control counts each sum to one across all clusters.

The marginal ratio of cases ($D$s) to negative controls ($\bar{D}$s) was 1:4, with 1000 cases and 4000 controls selected for each simulation. Five† intervention relative risks ($\lambda$ = 1, 0.8, 0.6, 0.4, 0.2) are examined and four different levels of differential ascertainment ($\alpha_{\mathrm{RA}}$ = 1, 0.95, 0.85, 0.5). The performance of the count ratio method of Section 2 ($\lambda_{\mathrm{R}}$) was compared with the bias-adjusted method of Section 3 ($\lambda_{\mathrm{TND}}$) using the variance estimates noted earlier.

For model-based comparisons we also consider mixed effects models and GEE. For the estimation of the relative risk using only case counts in the absence of a population-based denominator, the GEE and mixed effects models assume Poisson distributed counts and use

*There are two exceptions to the consecutive 2-year period counts. Data were missing in 2004 and 2009 which were ignored in making a 2-year time period in both cases.

†The supporting material also shows results for two additional intervention relative risks $\lambda$ = 0.5, 0.3.

a canonical log link. To estimate the relative risk with the inclusion of negative controls counts, the GEE and mixed effects models assume binomially distributed counts and use a canonical logit link. All mixed effects models include a random intercept for each cluster and all GEEs assume an exchangeable correlation structure.

All simulations and subsequent analyses were performed in R version 3.6.1 “Action of the Toes”.[14] GEE models were fit using “geeglm” from the “geepack” package.[15-17] Mixed effects models used “glmer” from the “lme4” package.[18] Plots were generated using the “ggplot2” package.[19] All additional simulation code is available as a GitHub repository managed by the first author.‡

## 5 | RESULTS

### 5.1 | Detecting an intervention effect

Figures 1 to 3 compare the performance of the count ratio estimator ($\lambda_R$) to the simple debiased estimator ($\lambda_{TND}$), as well as the mixed effects, and GEE approaches. The simulation results are averaged across the 10,000 unique intervention allocations applied to each of the nine different observed historical time periods. Thus, the simulations reflect overall performance over nine somewhat different scenarios. These results are summarized numerically in Tables S1 to S3 included in the Supporting Material.

Power, shown in Figure 1, is estimated as the proportion of permuted allocations that return a significant test result at a significance level of .05. Significance for the count ratio method is determined on the basis of the test statistic proposed in Equation (1), standardized by its estimated variance, compared with a $t$-distribution with $2(m-1)$ degrees of freedom. In the case of the simple ascertainment debiased estimator ($\lambda_{TND}$), a significant result is determined by the absence of the null value in the 95% confidence interval around the estimated intervention RR, as performed on the log scale. Finally, significance is determined by the model-based coefficient $P$-value corresponding to intervention in the mixed effects and GEE models. The power for each intervention and differential ascertainment scenario is relatively stable for the approaches that make use of both count and negative control information (Figure 1B). The count-only approach shows the most desirable estimated type I error in the setting where there is no differential ascertainment (power = 0.058). However, it seriously deteriorates for a high level of differential ascertainment. This is explained by the introduced bias in estimation. This does not affect the approaches that use the negative control information (Figure 1B), although there is some anticonservativeness in the simple TND estimator for a high level of differential ascertainment. The increasing power of the count-only methods (Figure 1A) for any fixed value of $\lambda$ is an artifact of the fact that, for the simulations considered here, the intervention effect and the differential ascertainment work in the same direction (of reducing counts in the intervention clusters); for simulations with $a_{RA} > 1$ (not shown here), the power of the count-only approaches substantially worsens as differential ascertainment widens.

‡https://github.com/sdufault15/case-only-crtnd

Bias (Figure 2) is estimated as $E_p[\hat{\lambda}] - \lambda$. The estimated bias is reported on the scale of the relative risk for inter-pretability. In the setting of no differential ascertainment ($a_{RA} = 1$), the estimators perform similarly, as expected, as most of the estimators enjoy zero asymptotic bias (note that the mixed effects model estimates a cluster-specific odds ratio that is not identical to the marginal odds ratio targeted by GEE and the other methods). The small gain when using the count ratio estimator is less than 1% which is negligible. Furthermore, as differential ascertainment increases, the count-only estimators (Figure 2A) are unable to reliably estimate the intervention effect. The simple TND estimator, binomial GEE, and binomial mixed effects methods (Figure 2B) all maintain low bias (bias ≤ 0.05).

Finally, coverage (Figure 3) represents the proportion of estimated 95% confidence intervals, which contain the true intervention relative risk. Again, in the absence of differential ascertainment, the count ratio estimator ($\lambda_R$) enjoys slightly improved coverage across each of the examined intervention RRs (≈ 93.4% coverage). As expected, however, the coverage deteriorates as the bias from differential ascertainment increases (Figure 3A). Slight deterioration in coverage as differential ascertainment worsens was observed across each of the estimators, though for the approaches accounting for negative controls (Figure 3B) coverage fell only to 90%.

### 5.2 | Detecting differential ascertainment

As described in Section 3, the count ratio estimator can be used to estimate the relative risk of differential ascertainment ($a_{RA}$) using the negative control counts, when available. Table 2 presents the bias, power, and coverage statistics for estimation of $a_{RA}$ when the true $a_{RA}$ is null ($a_{RA} = 1$), low ($a_{RA} = 0.95$), medium ($a_{RA} = 0.85$), and high ($a_{RA} = 0.5$). As the distribution of the negative controls is assumed unaffected by the intervention, these results are true for any size of intervention effect $\lambda$. Despite the low bias in estimation, good coverage, and type I error (ie, power when $a_{RA} = 1$), the power to detect differential ascertainment away from the null (ie, $a_{RA} \neq 1$) is necessarily low except with high differential ascertainment.

Note that, in Table 2, when $a_{RA} = 1$ (ie, at the null of no differential ascertainment), the power represents the Type I error and should be complementary to the coverage rate in that the two values should sum to 1. However, for the count ratio estimator, hypothesis testing is based on the normalized $t$ statistic of Section 3, whereas coverage is based on the confidence interval associated with the ratio estimator of the relative ascertainment $a_{RA}$, also introduced in Section 3. Thus, the corresponding entries only approximately add to one.

## 6 | CONCLUSIONS

The count-only approaches for CRTs perform comparably in estimation of an intervention relative risk compared with alternatives that use additional negative control information (albeit at reduced power), but only in the absence of differential ascertainment. The count-only methods have reasonable bias and coverage properties (near 94%) and comparable power while maintaining a desirable type I error rate. These properties depend entirely on randomization and so cannot be used directly when the clusters are not randomized.

Furthermore, the performance of the count-only approaches falter in the presence of even relatively low-differential ascertainment ($a_{RA}$ = 0.95) as demonstrated by increases in bias and decreases in coverage. By contrast, methods that adjust for differential ascertainment by incorporating proposed negative control counts maintain desirable performance even under major differential ascertainment. Thus, the count-only estimators should only be used when there is no other alternative (despite this being currently standard) and should be treated with considerable caution if there is any possibility of differential ascertainment. The use of negative controls in CRTs provides an attractive option to remove, or reduce, the effect of differential ascertainment and should be used more widely.

Potentially, the results have more significance when considering stepped wedge designs rather than the parallel arm scenario considered here. Currently, almost all stepped wedge studies only consider an outcome of interest and do not employ negative controls to remove bias. Analytical results for the stepped wedge design in this context will be provided elsewhere.

Finally, determining whether differential ascertainment exists by the estimation approach proposed here is informative but lacks sufficient power to detect moderate differences by intervention arm. As such, determining whether a setting is appropriate for future estimation by the count-only approach will likely return uninformative results unless ascertainment is exceptionally differential ($a_{\mathrm{RA}} \leq 0.5$ or $a_{\mathrm{RA}} \geq 2.0$).

### 6.1 | Recommendations

The findings suggest two key recommendations. First, in CRTs where only counts are available for analysis, the proposed estimator is a viable option with desirable statistical properties. However, even with randomized interventions, it is only appropriately employed in settings where there is little to no differential ascertainment by intervention arm. This is likely most plausible under blinded intervention assignment. Second, in CRTs where differential ascertainment is likely or inevitable, negative control data are important for validity.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS

The authors acknowledge funding from the National Institute of Allergy and Infectious Diseases grant R56AI134724.

Funding information

National Institute of Allergy and Infectious Diseases, R56AI134724

## Abbreviations:

| | |
|---|---|
| **CR-TND** | cluster-randomized test-negative design |
| **GEE** | generalized estimating equations |

| | |
|---|---|
| **OFI** | other febrile illness |
| **RA** | relative ascertainment |
| **RR** | relative risk |
| **TND** | test-negative design |
| **TP** | test-positive |

## REFERENCES


1. Hayes RJ, Moulton LH. Cluster Randomised Trials. 2nd ed. New York, NY: Chapman and Hall/CRC Press; 2017.
2. Cornfield J Randomization by group: a formal analysis. Am J Epidemiol. 1978;108(2):100–102. [PubMed: 707470]
3. Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. Int J Epidemiol. 1999;28(2):319–326. [PubMed: 10342698]
4. Lipsitch M, Tchetgen T, Eric CT. Negative controls: a tool for detcting confounding and bias in observational studies. Epidemiology. 2010;21(3):383–388. [PubMed: 20335814]
5. Anders KL, Cutcher Z, Kleinschmidt I, et al. Cluster-randomized test-negative design trials: a novel and efficient method to assess the efficacy of community-level dengue interventions. Am J Epidemiol. 2018;187(9):2021–2028. [PubMed: 29741576]
6. Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical basis of the test-negative study design for assessment of influenza vaccine effectiveness. Am J Epidemiol. 2016;184(5):345–353. [PubMed: 27587721]
7. Jackson ML, Nelson JC. The test-negative design for estimating influenza vaccine effectiveness. Vaccine. 2013;31(17):2165–2168. [PubMed: 23499601]
8. Johnson K The impact of Wolbachia on virus infection in mosquitoes. Viruses. 2015;7(11):5705–5717. 10.3390/v7112903. [PubMed: 26556361]
9. Jewell NP, Dufault S, Cutcher Z, Simmons CP, Anders KL. Analysis of cluster-randomized test-negative designs: cluster-level methods. Biostatistics. 2019;20(2):332–346. [PubMed: 29447357]
10. Anders KL, Indriani C, Ahmad RA, et al. The AWED trial (Applying Wolbachia to Eliminate Dengue) to assess the efficacy of Wolbachia-infected mosquito deployments to reduce dengue incidence in Yogyakarta, Indonesia: study protocol for a cluster randomised controlled trial. Trials. 2018;19(1):1–16. 10.1186/s13063-018-2670-z. [PubMed: 29298706]
11. Westreich D, Hudgens MG. Invited commentary: beware the test-negative design. Am J Epidemiol. 2016;184(5):354–356. [PubMed: 27587722]
12. Ferdinands JM, Foppa IM, Fry AM, Flannery BL, Belongia EA, Jackson ML. Re:"Invited commentary: beware the test-negative design". Am J Epidemiol. 2017;185(7):613–613. [PubMed: 28338844]
13. Chua H, Feng S, Lewnard JA, et al. The use of test-negative controls to monitor vaccine effectiveness. Epidemiology. 2020;31(1):43–64.
14. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.
15. Højsgaard S, Halekoh U, Yan J. The R package geepack for generalized estimating equations. J Stat Softw. 2006;15(2):1–11.
16. Yan J, Fine J. Estimating equations for association structures. Stat Med. 2004;23:859–880. [PubMed: 15027075]
17. Yan J Geepack: yet another package for generalized estimating equations. R-News. 2002;2(/3):12-14.
18. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1–48. 10.18637/jss.v067.i01.

19. Wickham H Ggplot2: Elegant Graphics for Data Analysis. New York, NY: Springer-Verlag; 2016.

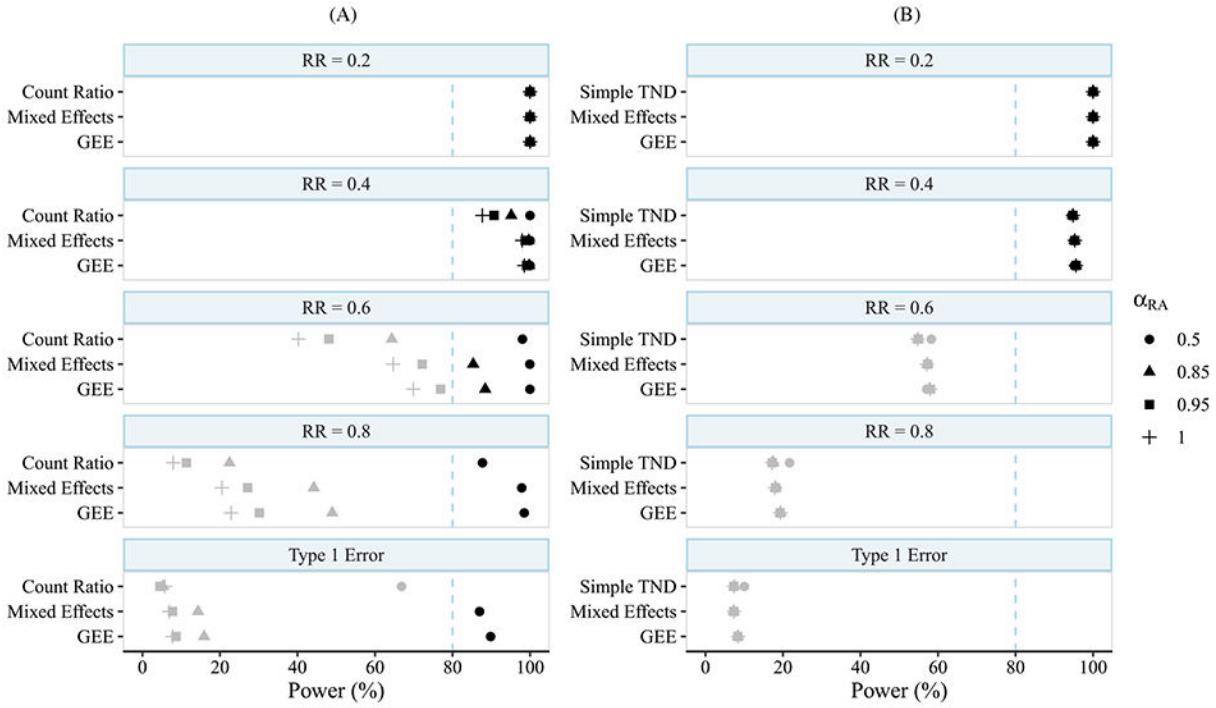


**FIGURE 1.**

The power, and Type I error rates, in testing departure from the null of no intervention effect based on various estimation methods for a range of relative risks (RR), over 10,000 intervention allocations applied to each of nine historical time periods with 1000 cases and 4000 negative controls (when applicable). Differential ascertainment ($\alpha_{RA}$) is allowed to increase in severity. A, Results from count-only methods in the absence of a population denominator. The mixed effects and generalized estimating equations (GEE) models assume the case counts are Poisson distributed and use the canonical log link. B, Negative control bias-adjusted results. The mixed effects and GEE models assume the case and negative control counts are binomially distributed and use the canonical logit link

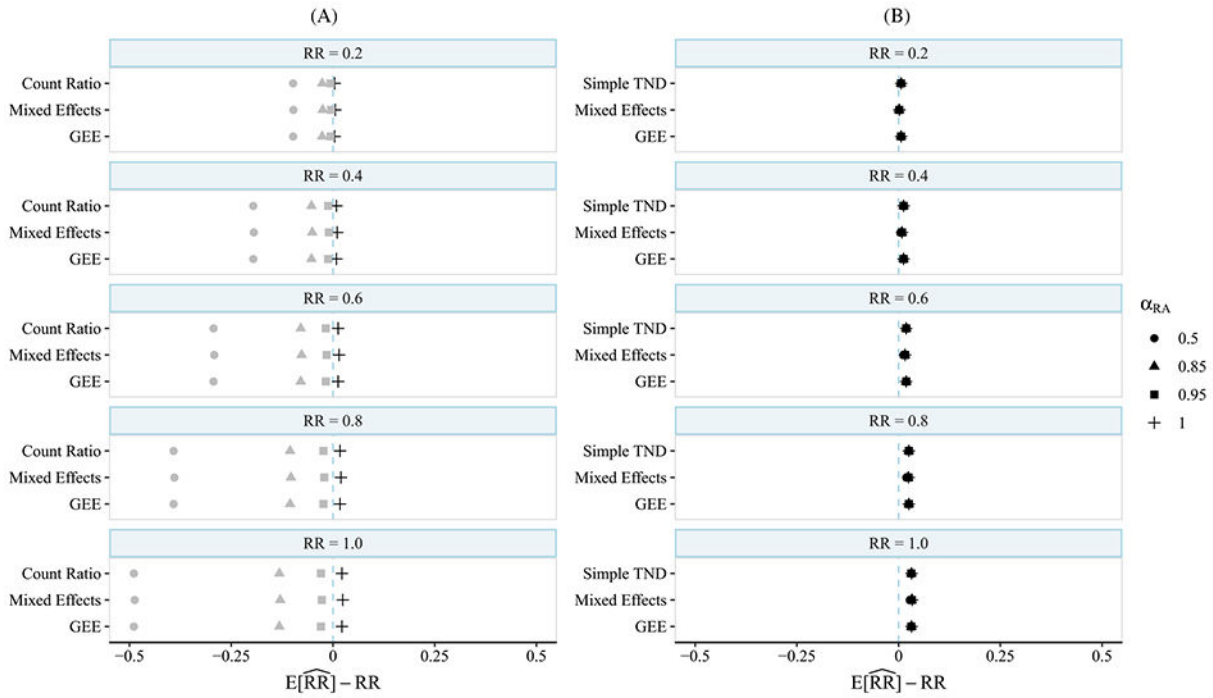


**FIGURE 2.**

Bias in estimation of the intervention relative risk (RR) for various methods over 10,000 intervention allocations applied to each of nine historical time periods with 1000 cases and 4000 negative controls as differential ascertainment increases in severity. A, Results from count-only methods in the absence of a population denominator. The mixed effects and generalized estimating equations (GEE) models assume the case counts are Poisson distributed and use the canonical log link. B, Negative control bias-adjusted results. The mixed effects and GEE models assume the case and negative control counts are binomially distributed and use the canonical logit link

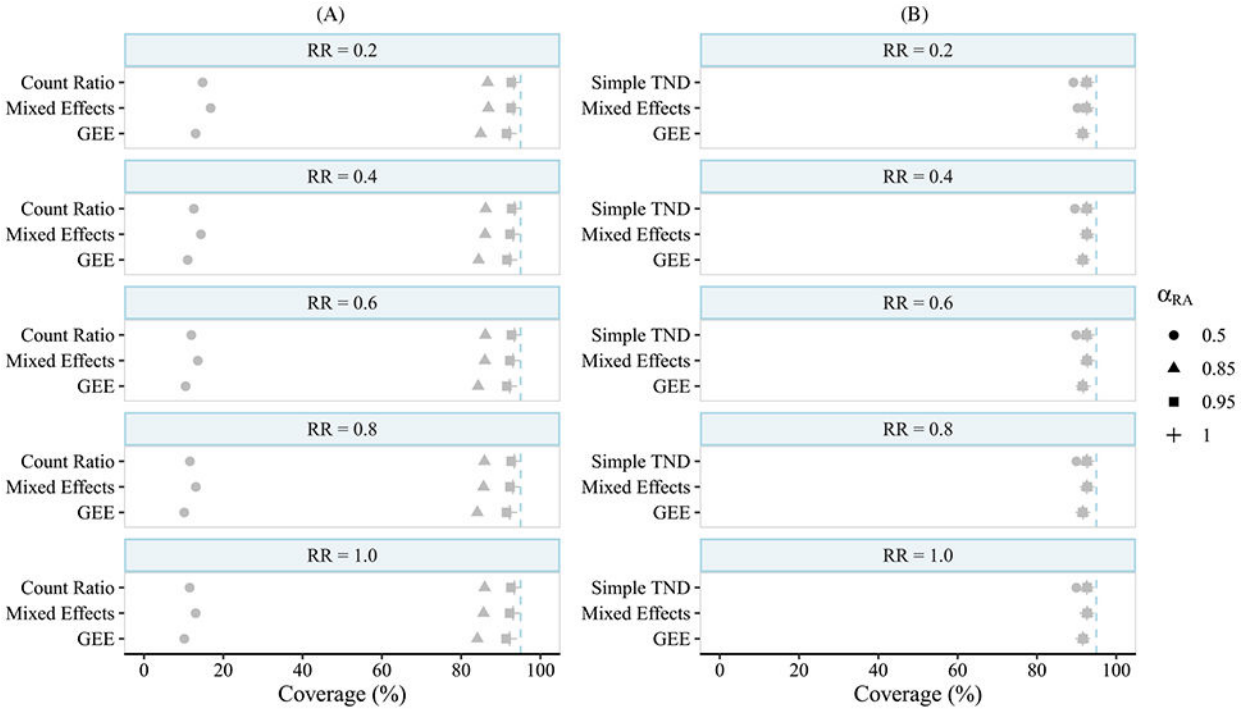


**FIGURE 3.**

95% confidence interval coverage based on estimation of the intervention relative risk (RR) for various methods over 10,000 intervention allocations applied to each of nine historical time periods with 1000 cases and 4000 negative controls as differential ascertainment increases in severity A, Results from the comparison of counts in the absence of population denominator. The mixed effects and generalized estimating equations (GEE) models assume the case counts are Poisson distributed and use the canonical log link. B, Bias-adjusted results. The mixed effects and GEE models assume the case and negative control counts are binomially distributed and use the canonical logit link

**TABLE 1**

Stratification of population based on intervention status, infection, and health-care-seeking behavior

| | Seek care | | | | Do not seek care | | | |
|---|---|---|---|---|---|---|---|---|
| | **Test-positive cases** | **Test-negative controls** | **Not infected** | **Total** | **Test-positive cases** | **Test-negative controls** | **Not infected** | **Total** |
| Intervention ($E$) | $A_T$ | $B_T$ | $C_T$ | $N_{IO}$ | $D_T$ | $E_T$ | $F_T$ | $N_{IU}$ |
| Control ($\bar{E}$) | $G_T$ | $H_T$ | $I_T$ | $N_{CO}$ | $J_T$ | $K_T$ | $L_T$ | $N_{CU}$ |

Note: Adapted from figure 1 of Jackson and Nelson.[7]

**TABLE 2**

Bias, power (type I error when $\alpha_{RA} = 1$), and 95% confidence interval coverage based on estimation of differential ascertainment by intervention arm from 10,000 permuted intervention allocations across nine time periods of historical data for a ratio of 1000 cases to 4000 negative controls

| | Bias | Power | Coverage |
|---|---|---|---|
| $\alpha_{RA} = 1$ | 0.0215 | 0.0583 | 0.935 |
| $\alpha_{RA} = 0.95$ | 0.0207 | 0.0479 | 0.934 |
| $\alpha_{RA} = 0.85$ | 0.0181 | 0.0443 | 0.935 |
| $\alpha_{RA} = 0.5$ | 0.0108 | 0.6870 | 0.934 |