Sample size calculation for cluster randomized trials with zero-inflated count outcomes

Zhengyang Zhou1 | Dateng Li2 | Song Zhang3

1 Department of Biostatistics and Epidemiology, University of North Texas Health Science Center, Fort Worth, Texas, USA
2 Early Clinical Development, Biostatistics, Regeneron Pharmaceuticals Inc., Tarrytown, New York, USA
3 Department of Population and Data Sciences, University of Texas Southwestern Medical Center, Dallas, Texas, USA

Correspondence
Dateng Li, Early Clinical Development, Biostatistics, Regeneron Pharmaceuticals Inc., Tarrytown, NY, USA.
Email: lidateng2015@gmail.com

Abstract
Cluster randomized trials (CRT) have been widely employed in medical and public health research. Many clinical count outcomes, such as the number of falls in nursing homes, exhibit excessive zero values. In the presence of zero inflation, traditional power analysis methods for count data based on Poisson or negative binomial distribution may be inadequate. In this study, we present a sample size method for CRTs with zero-inflated count outcomes. It is developed based on GEE regression directly modeling the marginal mean of a zero-inflated Poisson outcome, which avoids the challenge of testing two intervention effects under traditional modeling approaches. A closed-form sample size formula is derived which properly accounts for zero inflation, ICCs due to clustering, unbalanced randomization, and variability in cluster size. Robust approaches, including t-distribution-based approximation and Jackknife re-sampling variance estimator, are employed to enhance trial properties under small sample sizes. Extensive simulations are conducted to evaluate the performance of the proposed method. An application example is presented in a real clinical trial setting.

Keywords
cluster randomized trials, generalized estimating equation, marginalized models, sample size, zero-inflated outcomes

1 | INTRODUCTION

Clinical trials that perform randomization at the cluster level (eg, clinics, schools, communities, etc.) have been widely used to assess effectiveness of interventions in medical and public health research. Such trials are commonly referred to as cluster randomized trials (CRTs). Since participants in the same cluster share certain characteristics (eg, the same physician, teacher, similar socioeconomic status, etc.), their responses tend to be positively correlated. This intraclass correlation coefficient (ICC) is one of the key features that need to be considered in sample size calculation for CRTs.

Count outcomes are frequently used in CRTs. Examples include number of cigarettes in a smoking cessation study, number of clinic visits in an educational outreach study, and number of days in ICU in a nutritional support study. To model count data, the Poisson distribution has been widely used. Many researchers, however, have reported the phenomenon of zero inflation, where the observed proportion of zeros is much greater than the theoretical proportion under Poisson. For example, in a long-term care (LTC) RCT evaluating the effect of multifaceted knowledge translation for care teams, one outcome was the number of falls over three months among senior residents in LTC homes. This outcome
exhibited substantial zero-inflation where more than 30% of residents did not have any fall in either the control or intervention group during the follow-up period. Furthermore, zero-inflated count data often manifests overdispersion, which violates the assumption of mean and variance being equal under the Poisson model. Imposing the Poisson assumption on zero-inflated count outcomes leads to poor statistical results including biased estimation and under-estimated sample sizes. Recently, the negative binomial (NB) model has gained popularity due to its flexibility in accommodating overdispersion. The NB model, however, does not address the issue of zero-inflation either.

Most of existing sample size methods for CRTs with count outcomes are developed under the Poisson model. For example, Amatya et al proposed a sample size calculation method based on Poisson regression. It required the assumption of equal cluster size, which might be unrealistic in real-world clinical settings. It has been shown that ignoring variability in cluster size leads to under-powered studies. Wang et al relaxed this assumption and proposed a sample size method accommodating randomly varying cluster sizes. It included a correction term involving a coefficient of variance for cluster sizes. Li et al proposed a sample size method for correlated count outcome based on the NB model. Developed based on either Poisson or NB, the aforementioned methods are inapplicable to CRTs where the count outcomes contain excessive zeros.

The zero-inflated Poisson (ZIP) model has been widely used to analyze count data with excessive zeros. It assumes the count variable to arise from a mixture of a Poisson distribution and a point mass at zero (ie, the structural zero). This mixture distribution is characterized by two parameters: the Poisson mean and the probability of structural zero. Separate hypothesis testings can be performed on these two parameters, which assess the intervention effect in two dimensions. Notably, under the ZIP model, the intervention effects are estimated based on two latent portions. The Poisson portion represents a susceptible subpopulation of individuals who are at risk for an outcome of interest, and the structural zero portion represents a non-susceptible subpopulation of individuals who are no longer at risk. Such interpretation are useful in separating the intervention effect into two dimensions. However, it is cumbersome to synthesize the results to obtain the overall intervention effect (ie, difference in mean between the intervention and control groups), which has been the main research interest in many clinical trials (eg, References 4,12,22). Based on the ZIP model, power analysis and sample size calculation targeting the Poisson portion and the structural zero portion separately have been studied for independent data (eg, individual randomized trials). Recently, an extension to clustered data (eg, CRTs) has been discussed based on the hierarchical ZIP model with cluster-specific random effects. However, none of the existing methods provide solution of sample size calculation based on the power targeted for the overall intervention effect.

To address the above challenges, researchers have proposed the marginalized ZIP model, which provides a way to obtain an overall intervention effect. Instead of evaluating the two intervention effects, this approach directly makes inference on the overall intervention effect, quantified by the marginal mean under the ZIP framework. Compared to the ZIP model, this approach aims to assess impact of intervention on the target population as a whole. Other developments related to the marginal means of outcomes with excessive zeros include extensions to other types of distributions, extensions to clustered data by incorporating random effects, model selections methods, and statistical software to fit marginalized ZIP models. In this study, we build upon this idea to develop power analysis methods for the comparison of overall intervention effect in CRTs with zero-inflated count outcomes. The resulting sample size formula has a closed form, which facilitates implementation and enables researchers to analytically evaluate the impact of various design parameters. It also accommodates pragmatic design issues frequently encountered by practitioners such as unbalanced randomization and randomly varying cluster sizes. In addition, although the study focuses on power analysis of CRTs, the proposed method can be readily adapted to individual randomized trials.

The rest of the article is organized as follows. In Section 2, we describe the statistical model and power analysis approach for CRTs with zero-inflated outcomes. In Section 3, we conduct extensive simulations to evaluate the performance of the proposed method. In Section 4, a real application example is presented. In Section 5, we provide discussion and concluding remarks.

2 | METHODOLOGY

2.1 | Statistical model and sample size

Suppose $N$ clusters are randomized to the control or intervention arm in a CRT. We use $m_i$ ($i = 1, \ldots, N$) to denote the cluster sizes and $m_i$ are assumed to follow a certain discrete distribution: $\text{Prob}(m_i = m) = g(m)$ with outcome space $\mathcal{M}$. 
We define mean $\eta_m = E(m_i)$ and variance $\sigma^2_m = \text{Var}(m_i)$. Let $y_{ij}$ be the count outcome measured on the $j$th subject from the $i$th cluster. We assume that $y_{ij}$ arises from a ZIP distribution, which is the mixture of two components: a point mass at zero with probability $p_{ij}$, and a Poisson distribution of mean $\lambda_{ij}$ with probability $1 - p_{ij}$. Presented through latent variables, we have

$$y_{ij} = \begin{cases} 0 & \text{if } s_{ij} = 1; \\ u_{ij} & \text{if } s_{ij} = 0, \end{cases}$$

where $s_{ij}$ is binary with $\text{Prob}(s_{ij} = 1) = p_{ij}$ and $u_{ij} \sim \text{Poisson}(\lambda_{ij})$. Here $s_{ij}$ and $u_{ij}$ are latent variables. We define ICCs $\rho_s = \text{Corr}(s_{ij}, s_{ij'})$ and $\rho_u = \text{Corr}(u_{ij}, u_{ij'})$ for $j \neq j'$. Responses are assumed to be independent across clusters. Following prior literature,33,34 we further assume $\text{Corr}(u_{ij}, s_{ij'}) = 0$ and $\text{Corr}(u_{ij}, s_{ij}) = 0$. Beckett et al35 showed that the marginal mean and variance of $y_{ij}$ are

$$E(y_{ij}) = \mu_{ij} = (1 - p_{ij})\lambda_{ij},$$

and

$$\text{Var}(y_{ij}) = \mu_{ij} + \frac{p_{ij}}{1 - p_{ij}}\mu^2_{ij}.$$  

It is obvious that $\text{Var}(y_{ij})$ is an increasing function of $\mu_{ij}$ and $p_{ij}$. Furthermore, $\text{Var}(y_{ij}) > E(y_{ij})$ always holds. That is, given the same mean, a ZIP variable has a larger variance than a Poisson variable. Hence zero inflation implies over-dispersion. Mis-specifying a Poisson model for a ZIP variable would lead to underestimated variability in data analysis, and under-powered clinical trials in experimental design. The severity of over-dispersion is associated with a larger mean ($\mu_{ij}$) and a larger probability of structural zero ($p_{ij}$).

Traditionally researchers have evaluated the intervention effect by testing two hypotheses, one constructed based on $p_{ij}$ and the other based on $\lambda_{ij}$. Such approaches lead to difficulty in sample size calculation because statistical inference involves testing two hypotheses. For CRTs, this difficulty is further complicated by the need to consider clustering.

In this study we propose to directly evaluate the overall intervention effect, measured on the marginal mean ($\mu_{ij}$) of a ZIP outcome. Specifically, we assume

$$\log(\mu_{ij}) = \beta_1 + \beta_2 r_i.$$  

Here $r_i = 0/1$ indicates that the $i$th cluster is randomized to the control/intervention arm and $\beta_2$ is the difference in marginal mean between the intervention and control group on the log scale, representing the overall intervention effect.26 A cluster receives intervention with probability $\bar{r} = E(r_i)$. Define $\mu_i = \exp(\beta_1 + \beta_2 r_i)$ and model (3) suggests that $\mu_{ij} = \mu_i$. Similarly, we assume $p_{ij} = p_i$ and $\lambda_{ij} = \lambda_i = \mu_i/(1 - p_i)$. The hypotheses of interest are $H_0 : \beta_2 = 0$ vs $H_1 : \beta_2 \neq 0$.

With Equations (3) and (2), models for the first two moments of $y_{ij}$ have been specified. We can estimate the regression parameters $\beta$ using the generalized estimating equation (GEE) approach.36 Define $Z_{ij} = (1, r_i)' = Z_i$ and $\beta = (\beta_1, \beta_2)'$. Let $y_i = (y_{i1}, \ldots, y_{im_i})'$ be the cluster-specific response vector with mean $\mu_i(\beta) = [\mu_{i1}(\beta), \ldots, \mu_{im_i}(\beta)]' = \mu_i \mathbb{1}_{m_i}$, where $\mu_i = \exp(Z_i' \beta)$ and $\mathbb{1}_{m_i}$ is a vector of length $m_i$ with all elements being 1. Utilizing the independent working correlation, the GEE estimator $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)'$ is the solution to score function

$$S_N(\beta) = N^{-\frac{1}{2}} \sum_{i=1}^N D_i' W_i^{-1} [y_i - \mu_i(\beta)] = 0,$$

where $D_i = \frac{\partial \mu_i(\beta)}{\partial \beta}$ is an $m_i \times 2$ gradient matrix and $W_i$ is an $m_i \times m_i$ diagonal matrix with all diagonal elements being $\mu_i + \frac{p_i}{1 - p_i} \mu^2_i$. Equation (4) can be solved through the Newton-Raphson algorithm. Specifically, at the $l$th iteration,

$$\hat{\beta}^{(l+1)} = \hat{\beta}^{(l)} + N^{-\frac{1}{2}} A_N^{-1}(\hat{\beta}^{(l)} - \beta) S_N(\hat{\beta}^{(l)}).$$
where

\[
A_N(\hat{\beta}) = N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{m_i} Z_i Z'_j \frac{\mu_i(\hat{\beta})}{1 + \frac{\rho_i}{1-p_i} \mu_i(\hat{\beta})}.
\] (6)

As shown by Liang and Zeger,\textsuperscript{36} \(\sqrt{N}(\hat{\beta} - \beta)\) approximately follows a normal distribution with mean 0 and variance \(\Sigma_N = A_N^{-1} V_N A_N^{-1}\), where

\[
V_N(\hat{\beta}) = N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{m_i} \sum_{p=1}^{m_p} \hat{e}_{ij} \hat{e}'_{ij} \frac{1}{ [1 + \frac{\rho_i}{1-p_i} \mu_i(\hat{\beta})]^2 } Z_i Z'_j.
\]

Here \(\hat{e}_{ij} = y_{ij} - \exp(Z'_i \hat{\beta})\) is the residual. Let \(\sigma_2^2\) be the (2,2)th element of \(\Sigma_N\). We reject \(H_0: \beta_2 = 0\) if \(\sqrt{N} |\hat{\beta}_2| / \sigma_2 > z_{1-\alpha/2}\), where \(z_{1-\alpha/2}\) is the 100(1 - \(\alpha/2\))th percentile of the standard normal distribution. Define \(A\) and \(V\) to be the limits of \(A_N\) and \(V_n\) as \(N \to \infty\). It follows that \(\Sigma_N\) converges to \(\Sigma = A^{-1} VA^{-1}\). Let \(\sigma_2^2\) be the (2,2)th element of \(\Sigma\). Given the true intervention effect \(\beta_2 = \beta_2^0\), the number of clusters to achieve power 1 - \(\gamma\) at two-sided type I error \(\alpha\) is calculated by

\[
N = \frac{\sigma_2^2 (z_{1-\alpha/2} + z_{1-\gamma})^2}{\beta_2^0}.
\] (7)

In the following we show that a closed-form expression of \(\sigma_2^2\) can be derived, which leads to a closed-form sample size formula. First, as \(N \to \infty\), it is easy to show that \(A_N\) approaches

\[
A = (1 - \bar{r}) \frac{\mu^*_1 \eta_m}{1 + \frac{\rho_i}{1-p_i} \mu^*_1} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \bar{r} \frac{\mu^*_2 \eta_m}{1 + \frac{\rho_i}{1-p_i} \mu^*_2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.
\] (8)

Here we define \(p_i = p_1^*\) to be the probability of structural zero under control \((r_i = 0)\) and \(p_i = p_2^*\) under intervention \((r_i = 1)\). Similarly, \(\mu^*_1 = \exp(\beta_1)\) and \(\mu^*_2 = \exp(\beta_1 + \beta_2)\) are the marginal means. Recall that \(\eta_m = E(m_i)\). As \(N \to \infty\), \(V_N(\hat{\beta})\) approaches

\[
V = E \left[ \sum_{j=1}^{m_i} \sum_{p=1}^{m_p} \frac{[y_{ij} - \mu_i(\hat{\beta})][y_{ij}' - \mu_i(\hat{\beta})]}{[1 + \frac{\rho_i}{1-p_i} \mu_i(\hat{\beta})]^2} \begin{pmatrix} 1 & r_i' \\ r_i & r_i' \end{pmatrix} \right].
\]

\[
= (1 - \bar{r}) V_1 + \bar{r} V_2,
\] (9)

where

\[
V_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \left\{ \frac{\mu^*_1 + \frac{\rho_i}{1-p_i} \mu^*_1}{1 + \frac{\rho_i}{1-p_i} \mu^*_1} \frac{\eta_m + \sigma_m - \eta_m}{1 + \frac{\rho_i}{1-p_i} \mu^*_1} \right\}.
\]

and

\[
V_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \left\{ \frac{\mu^*_2 + \frac{\rho_i}{1-p_i} \mu^*_2}{1 + \frac{\rho_i}{1-p_i} \mu^*_2} \frac{\eta_m + \sigma_m - \eta_m}{1 + \frac{\rho_i}{1-p_i} \mu^*_2} \right\}.
\]

The terms \(\zeta_1\) and \(\zeta_2\) have relatively complicated expressions,

\[
\zeta_1 = \mu^*_1 p_1^*(1 + p_1^*) \rho_2 + \mu^*_1 \rho_2 (1 - p_1^* + p_1^* \rho_2) + \frac{\mu^*_1 \rho_1^2 \rho_2}{1 - p_1^*},
\]
Decomposing the marginal treatment effect

Aim at controlling alcohol consumption tends to increase the proportion of abstainers as well. Hence

Here the degree of freedom for the samplesize formula: usually there is no closed-form formula for sample size calculation based on the $t$-distribution. Tang et al. proposed a two-step procedure to obtain sample size under the $t$-distribution:

\[
N^{(i)} = \frac{n_{m}n_{c}^{*} \left(1 + \frac{\mu_{1}^{*}}{1 - \rho_{1}} \mu_{1}^{*}\right) + \left(\eta_{m}^{*} + \sigma_{m}^{*} - \eta_{m}\right) \xi_{1}}{(1 - \rho)\mu_{1}^{*}^{2} \eta_{m}^{*}} + \frac{n_{m}n_{c} \left(1 + \frac{\mu_{2}^{*}}{1 - \rho_{2}} \mu_{2}^{*}\right) + \left(\eta_{m} + \sigma_{m}^{*} - \eta_{m}\right) \xi_{2}}{\rho \mu_{2}^{*}^{2} \eta_{m}^{*}}. \tag{11}
\]

Sample size $N^{(i)}$ is obtained under asymptotic normal approximation. In practice, when the number of clusters is limited, the normal approximation might not perform well. In such cases, an alternative approach is to use the $t$-distribution. Usually, there is no closed-form formula for sample size calculation based on the $t$-distribution. Tang et al. proposed a two-step procedure to obtain sample size under the $t$-distribution:

\[
N^{(i)} = \frac{n_{m}n_{c}^{*} \left(1 + \frac{\mu_{1}^{*}}{1 - \rho_{1}} \mu_{1}^{*}\right) + \left(\eta_{m}^{*} + \sigma_{m}^{*} - \eta_{m}\right) \xi_{1}}{(1 - \rho)\mu_{1}^{*}^{2} \eta_{m}^{*}} + \frac{n_{m}n_{c} \left(1 + \frac{\mu_{2}^{*}}{1 - \rho_{2}} \mu_{2}^{*}\right) + \left(\eta_{m} + \sigma_{m}^{*} - \eta_{m}\right) \xi_{2}}{\rho \mu_{2}^{*}^{2} \eta_{m}^{*}}. \tag{12}
\]

Here the degree of freedom for the $t$-distribution is computed by a function of $N^{(i)}$, denoted as $f(N^{(i)})$. We set $f(N^{(i)}) = N^{(i)} - 2$, which equals to the number of clusters minus the number of regression parameters.

In summary, to compute a sample size using (11) or (12), we need to specify the mean and variance of cluster sizes ($\eta_{m}, \sigma_{m}^{2}$), the randomization probability $\bar{r}$, the regression parameters $\beta$, the probabilities of structural zeros ($p_{1}^{*}, p_{2}^{*}$), the ICC parameters ($\rho_{1}, \rho_{2}$), and pre-determined levels of type I error $\alpha$ and power $1 - \gamma$. Furthermore, the proposed sample size formulas accommodate individual randomized trials as a special case, where every cluster is of size 1. This implies that $\eta_{m} = 1$ and $\sigma_{m} = 0$. It is then straightforward to show that $\sigma_{2}^{2} = \frac{(1 + \frac{\mu_{1}^{*}}{1 - \rho_{1}} \mu_{1}^{*})}{(1 - \rho_{1})\mu_{1}^{*}^{2} \eta_{m}^{*}} + \frac{(1 + \frac{\mu_{2}^{*}}{1 - \rho_{2}} \mu_{2}^{*})}{\rho \mu_{2}^{*}^{2} \eta_{m}^{*}}$ and the sample size formula can be obtained accordingly.

### 2.2 Decomposing the marginal treatment effect $\beta_{2}$

Following the definition of ($p_{1}^{*}, p_{2}^{*}$) and ($\mu_{1}^{*}, \mu_{2}^{*}$), we define $\lambda_{1}^{*}$ and $\lambda_{2}^{*}$ to be the Poisson mean under control and intervention, respectively. Hence $\mu_{k}^{*} = (1 - p_{k}^{*}) \lambda_{k}^{*}$ for $k = 1, 2$. From (3) we have

\[
\beta_{2} = \log(\mu_{2}^{*}) - \log(\mu_{1}^{*}) = \log(\lambda_{2}^{*}) - \log(\lambda_{1}^{*}) + \log(1 - p_{2}^{*}) - \log(1 - p_{1}^{*}).
\]

That is, the overall intervention effect $\beta_{2}$ can be decomposed into $\log(\lambda_{2}^{*}) - \log(\lambda_{1}^{*})$ and $\log(1 - p_{2}^{*}) - \log(1 - p_{1}^{*})$, representing the effects on the Poisson part and the structural zero part, respectively. We introduce a new parameter $q$ such that

\[
\log(1 - p_{2}^{*}) - \log(1 - p_{1}^{*}) = q\beta_{2}. \tag{13}
\]

In practice, the intervention usually affects the Poisson part and the structural zero part in the same direction. For example, an intervention aimed at controlling alcohol consumption tends to increase the proportion of abstainers as well. Hence
we assume that \( \log(\lambda^*_2) - \log(\lambda^*_1) \) and \( \log(1 - p^*_2) - \log(1 - p^*_1) \) are of the same sign and \( q \in [0, 1] \). We interpret \( q \) as the proportion of treatment effect due to change in the probability of structural zeros.

It is straightforward that \( p^*_2 = 1 - \exp(q \beta_2)(1 - p^*_1) \). Hence during sample size calculation, we can equivalently specify either \((\beta_1, \beta_2, p^*_1, p^*_2)\), or \((\beta_1, \beta_2, p^*_1, q)\). We prefer the latter because it offers a straightforward decomposition of the overall intervention effect and a natural framework for sensitivity analysis. At the design stage, it is relatively easy to specify \((p^*_1, \beta_1)\) based on historical data, and \( \beta_2 \) based on what is considered a clinically meaningful change in the marginal mean. The specification of \( p^*_2 \), which is the probability of a latent variable, is difficult due to lack of information on the experimental intervention. With greater interpretability of \( q \), it is easier to solicit input from clinical experts. Furthermore, sensitivity analysis that explores a series of potential \( q \) values can be communicated back to clinicians as, taking the alcohol-controlling intervention for example, how sample size requirement varies with respect to the relative effect of the intervention on reducing a subject’s alcohol consumption vs transforming him/her into an abstiner.

Finally, given \( p^*_1 \), if \( \beta_2 > 0 \), a larger \( q \) is associated with a smaller \( p^*_2 \), and in turn a smaller variance \( \text{Var}(y_{ij}) \) under intervention \((n_1 = 1)\) according to (2). The association is in the opposite direction if \( \beta_2 < 0 \).

### 2.3 Estimating auxiliary parameters

The derivation of \( N^{(1)} \) and \( N^{(0)} \) in Section 2.1 assumes \((p^*_1, p^*_2)\) to be known. In actual data analysis they are most likely unknown and need to be estimated. Conventionally, researchers have modeled \( p_{ij} \) by a logit model,

\[
\log \left( \frac{p_{ij}}{1 - p_{ij}} \right) = \alpha_1 + \alpha_2 r_i. \quad (14)
\]

Statistically speaking, the parameterization by \( \alpha = (\alpha_1, \alpha_2)' \) is equivalent to that by \((p^*_1, p^*_2)\), with \( p^*_1 = \exp(\alpha_1)/(1 + \exp(\alpha_1)) \) and \( p^*_2 = \exp(\alpha_1 + \alpha_2)/(1 + \exp(\alpha_1 + \alpha_2)) \). Modeling approaches such as (14), however, offer greater flexibility to account for additional covariates. In the following we describe how to obtain \((\hat{p}^*_1, \hat{p}^*_2)\) through the estimation of \( \alpha \) using the expectation-solution algorithm.\(^{34}\)

First note that, if \( s_{ij} \) \((i = 1, \ldots, N; j = 1, \ldots, m_i)\) were observed, parameters \( \alpha \) could be estimated by solving a GEE equation:

\[
N^{-\frac{1}{2}} \sum_{i=1}^{N} \left[ \frac{\partial p_{i}(\alpha)}{\partial \alpha} \right] \left[ U^{-1}_i \right] [s_{i} - p_{i}(\alpha)] = 0. \quad (15)
\]

Here \( p_{i}(\alpha) = p_{i} \mathbb{1}_{m_i}, s_{i} = (s_{i1}, \ldots, s_{im_i})' \), and \( U_i \) is a \( m_i \times m_i \) diagonal matrix with all diagonal elements being \( p_i(1 - p_i) \).

Since \( s_{ij} \) is not observed when \( y_{ij} = 0 \), the solutions offered by (15) is not directly applicable. Instead, the expectation-solution algorithm can be employed which replaces \( s_{ij} \) in (15) with \( d_{ij} \), its conditional mean given \( \{y, \beta, \alpha\} \). Specifically,

\[
d_{ij} = \text{Prob}(s_{ij} = 1 | y, \beta, \alpha),
\]

\[
= \left\{ \frac{\text{Prob}(s_{ij} = 1, y_{ij} = 0 | y, \beta, \alpha)}{\text{Prob}(y_{ij} = 0 | y, \beta, \alpha)} \right\} I_{(y, = 0)},
\]

\[
= \left\{ 1 + \frac{1 - p_{ij} \exp(-\lambda_{ij})}{p_{ij}} \right\}^{-1} I_{(y, = 0)},
\]

\[
= \frac{p_{i}}{p_i + (1 - p_i) \exp(\lambda_i)} I_{(y, = 0)}. \quad (16)
\]

Defining \( d_i = (d_{i1}, \ldots, d_{im_i})' \), we modify (15) to

\[
N^{-\frac{1}{2}} \sum_{i=1}^{N} \left[ \frac{\partial p_{i}(\alpha)}{\partial \alpha} \right]' \left[ U^{-1}_i \right] [d_{i} - p_{i}(\alpha)] = 0. \quad (17)
\]
Finally, the complete estimation procedure is:

1. Obtain $\hat{\alpha}^{(0)}$, the initial value, by running a logistic regression using $\{I(y_i=0)\}$ as the response variable.
2. Plug $\hat{\alpha}^{(0)}$ into (4) to obtain $\hat{\beta}^{(1)}$. Given $\{\hat{\beta}^{(1)}, \hat{\alpha}^{(0)}\}$, calculate $d_{ij}^{(1)}$’s using (16). Then plug $d_{ij}^{(1)}$’s into (17) to obtain $\hat{\alpha}^{(1)}$.
3. Repeat Step 2 until the estimators converge.

Note that in Equation (4), $\hat{p}_1^* = \exp(\hat{\alpha}_1)/[1 + \exp(\hat{\alpha}_1)]$ and $\hat{p}_2^* = \exp(\hat{\alpha}_1 + \hat{\alpha}_2)/[1 + \exp(\hat{\alpha}_1 + \hat{\alpha}_2)]$. Kong et al also showed that the expectation-solution algorithm can be employed to estimate $\rho_s$ and $\rho_u$. 24

### 2.4 Addressing the issue of small sample sizes

In CRTs, the number of clusters ($N$) is often limited. 38 In such cases the sandwich-type variance estimator $\Sigma_N$ is known to be biased downwards, leading to an inflated type I error. 39 Alternatively, $\sigma^2$ can be estimated using re-sampling based methods, such as the Jackknife approach. 40 Many researchers have shown that better inference results can be obtained using re-sampling methods. 41,42 Let $\Sigma_N^{(Jack)}$ denote the estimate of $\Sigma$ using the Jackknife approach. It is calculated by

$$\Sigma_N^{(Jack)} = \frac{N}{N} \sum_{i=1}^{N} (\hat{\beta}^{(-i)} - \hat{\beta})(\hat{\beta}^{(-i)} - \hat{\beta})',$$

where $\hat{\beta}^{(-i)}$ denotes the estimate of $\beta$ based on data excluding the $i$th cluster. Importantly, we perform the re-sampling step at cluster level instead of patient level, so that within-cluster correlation is preserved. 41 We denote the sandwich-type variance estimation approach as “GEE-Naive,” and the Jackknife approach by (18) as “GEE-Jackknife.”

### 3 SIMULATION

We conduct simulations to assess performance of the proposed sample size method in terms of empirical power and type I error. Suppose $N$ clusters are randomized 1:1 to the control and intervention arms ($\bar{t} = 0.5$). We assume cluster sizes ($m_i$) to be randomly varying, and three distributions are considered: a truncated Poisson distribution with a mean parameter 45 over a range of [20, 70], denoted by TrunPoisson(20,70) with mean $\eta_m \approx 45$ and variance $\sigma^2_m \approx 44.8$; a discrete uniform distribution (DU) over a range of [34,56], denoted by DU(34,56) with $\eta_m = 45$ and $\sigma^2_m = 44$; a DU(10,80) distribution with the same mean $\eta_m = 45$ but greater variability $\sigma^2_m = 420$. Define $\rho = (\rho_s, \rho_u)$, and two sets of ICCs are explored: $\rho = (0.03, 0.03), (0.05, 0.05)$. ICCs of similar magnitude have been frequently reported in CRTs. 3 The nominal levels of two-sided type I error and power are set at $\alpha = 0.05$ and $1 - \gamma = 0.8$, respectively. We set the regression parameters $\alpha_1 = \beta_1 = 0$, implying that for the control group, $y_{ij}$ has a 50% ($p_1^* = 0.5$) chance of being a structural zero, and the overall mean is $\mu_{ij} = 1$. The goal of the CRT is to assess whether the intervention reduces a count outcome (eg, the number of falls in nursing homes). We consider $\beta_2 = -0.431$ and $-0.511$, which corresponds to a decrease of 0.35 and 0.40 in the overall mean ($\mu_{ij}$ from 1 to 0.65 and 0.6), respectively. Finally, five values of $q$ are explored: $q = 0.3, 0.4, 0.5, 0.6, 0.7$, allowing a sensitivity analysis on the proportion of treatment effect due to change in the probability of structural zeros between groups. Given a particular combination of design configurations, the simulation scheme is described as follows:

1. Given parameters ($p_1^*, \beta_2, q$), compute the value of $p_2^*$.
2. Plug the design parameters into (7) to compute sample size $N^{(2)}$ or further plug $N^{(2)}$ into (12) to obtain $N^{(0)}$.
3. For each scenario, we run $L = 2000$ iterations. In the $l$th iteration,
   a. Generate a random dataset of $N^{(2)}$ or $N^{(0)}$ clusters under the alternative hypothesis ($\beta_2 = \beta_{20}$).
      i. For each cluster, first generate cluster size $m_i$ from the assumed TrunPoisson or DU distribution.
      ii. Randomize this cluster to control or intervention. Depending on which arm the cluster belongs to, the overall mean $\mu_i$ and the probability of being a structural zero $p_i$ are determined. Given $\mu_i$ and $p_i$, the mean of the Poisson part is determined (ie, $\lambda_i = \frac{\mu_i}{1-p_i}$).
iii. An \( m_1 \)-length vector of correlated binary variables \((s_{i1}, \ldots, s_{im})\) is generated using the method of Reference 43 with marginal probability \( p_i \) and ICC \( \rho_s \).

iv. An \( m_1 \)-length vector of correlated Poisson variables \((u_{i1}, \ldots, u_{im})\) is obtained by generating \( u_{ij} = v_{ij} + v_i' \), where \( v_i' \)'s are random variables from a Poisson distribution with mean \( \lambda_i(1 - \rho_u) \) and \( v_i' \) is generated from a Poisson distribution with mean \( \lambda_i \rho_u \).

v. Finally, the \( m_1 \)-length response vector \( y_i = (y_{i1}, \ldots, y_{im})' \) is obtained through operation \( y_{ij} = (1 - s_{ij})u_{ij} \) for \( j = 1, \ldots, m_i \).

(b) Based on the generated dataset, we obtain \( \hat{\beta}_2, \hat{\sigma}_2^{2(\text{Naive})}, \hat{\sigma}_2^{2(\text{Jack})} \), respectively.

4. Empirical power of the “GEE-Naive” approach is computed as the proportion of iterations where \( |\sqrt{N} \frac{\hat{\beta}_2}{\hat{\sigma}_2^{(\text{Naive})}}| > z_{1-0.05/2} \) for \( N^{(G)} \) and \( |\sqrt{N} \frac{\hat{\beta}_2}{\hat{\sigma}_2^{(\text{Naive})}}| > t_{N^{(G)}-2,1-0.05/2} \) for \( N^{(I)} \). The empirical powers of the “GEE-Jackknife” approach are computed using \( \hat{\sigma}_2^{(\text{Jack})} \).

5. Empirical type I error is obtained similarly except for setting \( \beta_2 = 0 \) in Step 3(a).

Tables 1 and 2 present the numbers of clusters \( N^{(G)} \) and \( N^{(I)} \), empirical type I error, and empirical power for the “GEE-Naive” and “GEE-Jackknife” approaches under different combinations of design parameters for \( \beta_2 = -0.431 \). Across all scenarios, the numbers of clusters \( N^{(G)} \) and \( N^{(I)} \) range from 18 to 30 and 21 to 32, respectively, with \( N^{(I)} \) being slightly larger than \( N^{(G)} \) in every setting. When the “GEE-Naive” approach is paired with \( N^{(G)} \), the empirical type I error and power tend to be seriously inflated. Pairing the “GEE-Naive” approach with \( N^{(I)} \), or the “GEE-Jackknife” approach with \( N^{(G)} \), leads to slightly better performance, but moderate inflation in type I error persists. Finally, the combination of the “GEE-Jackknife” approach and \( N^{(I)} \) achieves the best performance, with both type I error and power controlled at nominal levels across all scenarios. Therefore, in practice we recommend calculating sample size using the \( t \)-distribution-based formula and perform data analysis using the “GEE-Jackknife” approach. The results for a larger effect size of \( \beta_2 = -0.511 \) under \( N^{(I)} \) are reported in Table S1 in the Supplementary Materials, which shows that the combination of \( N^{(I)} \) sample size and the “GEE-Jackknife” approach also performs well for smaller sample sizes (ranging from 16 to 24).

Tables 1 and 2 show a monotone relationship between \( q \) and sample size, which is consistent with the theoretical property that if \( \beta_2 < 0 \), a larger \( q \) is associated with a larger variance \( \text{Var}(y_{ij}) \) under intervention. We also observe that the proposed sample size is quite robust to change in \( q \). In particular, for the sample sizes based on \( t \)-distribution approximation (ie, Table 2), the sample sizes are mostly within 5% from the median (obtained at \( q = 0.5 \)) as \( q \) varies from 0.3 to 0.7. The variation in sample size is smaller under weaker ICCs, that is, \((\rho_s, \rho_u) = (0.03, 0.03)\). The above observation suggests that in practice, when there is limited prior knowledge, using \( q = 0.5 \) as a default specification might provide a reasonable initial sample size assessment.

The proposed sample size method accommodates random varying cluster sizes through the mean and variance parameters \((\eta_m, \sigma_m^2)\). In Tables 1 and 2, the comparison of results between TrunPoisson (20, 70) and DU(34, 56) represents a sensitivity analysis on cluster size distributions, where the means and variances are comparable but the distributions are different. No significant difference is observed in simulation results between the distributions, suggesting that the proposed sample size method is robust to randomly varying cluster sizes of different distributions. On the other hand, the comparison between DU(34, 56) and DU(10, 80) represents another sensitivity analysis where the cluster sizes follow the same type of distribution (DU) with a common center \((\eta_m = 45)\), but the variances are different. The results show that larger variability in cluster size leads to larger sample size requirement.

To demonstrate the consequence of misusing Poisson-based power analysis methods for CRTs with ZIP outcomes, we compare the number of clusters calculated between the proposed method and the Poisson-based approach by Wang et al., which also accounts for random variability in cluster size. Since Reference 17 only considered normal distribution, we use Equation (11), that is, \( N^{(G)} \), to calculate the number of clusters for a fair comparison. For each configuration, we set the Poisson mean equal to the marginal mean of the ZIP model. Recall that ICCs are specified on both the Poisson part and structural zero part for the ZIP model. To obtain a comparable ICC for the Poisson model, we generate a ZIP data set with 10 000 clusters, and then estimate ICC by fitting a Poisson model with an “exchangeable” correlation structure using the GEE approach. We denote this estimated ICC as \( \hat{\rho}^{(\text{Poisson})} \). The resulting sample sizes are presented in Table 3. Mistakenly applying a Poisson-based method to a CRT with a zero-inflated outcome would lead to a severely under-powered clinical trial.
## Table 1
Simulation: Empirical type I error and power for $N^{(z)} (\beta_2 = -0.431)$

| (\rho_\text{xy}, \rho_u)     | $q$ | $N^{(z)}$ | GEE-Naive Type I error | Power | GEE-Jackknife Type I error | Power |
|-----------------------------|-----|-----------|------------------------|-------|---------------------------|-------|
| TrunPoisson (45,20,70)      | 0.3 | 18        | 0.085                  | 0.846 | 0.067                     | 0.814 |
|                             | 0.4 | 19        | 0.079                  | 0.858 | 0.062                     | 0.818 |
|                             | 0.5 | 19        | 0.080                  | 0.857 | 0.063                     | 0.821 |
|                             | 0.6 | 20        | 0.082                  | 0.851 | 0.065                     | 0.823 |
|                             | 0.7 | 20        | 0.082                  | 0.844 | 0.065                     | 0.815 |
|                             | (0.05,0.05) | 0.3    | 24        | 0.075      | 0.835                  | 0.061 | 0.807 |
|                             |     |           | 0.071                  | 0.849 | 0.061                     | 0.823 |
|                             |     |           | 0.071                  | 0.843 | 0.061                     | 0.818 |
|                             |     |           | 0.068                  | 0.849 | 0.055                     | 0.825 |
|                             |     |           | 0.068                  | 0.856 | 0.054                     | 0.827 |
| DU (34, 56)                 | 0.3 | 18        | 0.081                  | 0.848 | 0.063                     | 0.814 |
|                             | 0.4 | 19        | 0.083                  | 0.844 | 0.065                     | 0.817 |
|                             | 0.5 | 19        | 0.083                  | 0.845 | 0.065                     | 0.805 |
|                             | 0.6 | 20        | 0.086                  | 0.849 | 0.066                     | 0.823 |
|                             | 0.7 | 20        | 0.086                  | 0.848 | 0.066                     | 0.817 |
|                             | (0.05,0.05) | 0.3    | 24        | 0.077      | 0.836      | 0.061 | 0.805 |
|                             |     |           | 0.075                  | 0.852 | 0.063                     | 0.829 |
|                             |     |           | 0.075                  | 0.852 | 0.063                     | 0.820 |
|                             |     |           | 0.074                  | 0.843 | 0.062                     | 0.809 |
|                             |     |           | 0.071                  | 0.841 | 0.064                     | 0.818 |
| DU (10, 80)                 | 0.3 | 20        | 0.094                  | 0.851 | 0.075                     | 0.813 |
|                             | 0.4 | 20        | 0.095                  | 0.849 | 0.073                     | 0.804 |
|                             | 0.5 | 21        | 0.093                  | 0.859 | 0.069                     | 0.818 |
|                             | 0.6 | 21        | 0.093                  | 0.851 | 0.069                     | 0.803 |
|                             | 0.7 | 22        | 0.086                  | 0.847 | 0.069                     | 0.804 |
|                             | (0.05,0.05) | 0.3    | 27        | 0.074      | 0.828                  | 0.051 | 0.794 |
|                             |     |           | 0.073                  | 0.837 | 0.054                     | 0.802 |
|                             |     |           | 0.073                  | 0.829 | 0.054                     | 0.799 |
|                             |     |           | 0.067                  | 0.846 | 0.054                     | 0.811 |
|                             |     |           | 0.062                  | 0.835 | 0.049                     | 0.804 |

### 4 | Application

We apply the proposed sample size method to a CRT that evaluated the effectiveness of knowledge translation strategies for care team members in long-term care (LTC) settings. The outcome of interest was the number of falls for senior residents which exhibited zero-inflation. Forty LTC homes were included as clusters and the average number of participants per cluster was 137, with a range of [43, 375]. The study reported that, during a three month follow-up, the average number of falls and the proportion of zeros were (1.21, 37.2%) in the control arm.

Suppose we want to design a new CRT to investigate the effectiveness of a new intervention on reducing falls in LTC homes, where the power and two-sided type I error are set at 80% and 5%, respectively. The design parameters to estimate...
| (\(\rho_s, \rho_u\)) | \(q\) | \(N^{(t)}\) | Type I error | Power | Type I error | Power |
|-----------------|--------|-------------|--------------|--------|--------------|--------|
| TrunPoisson (45,20,70) | (0.03, 0.03) | 0.3 | 21 | 0.066 | 0.862 | 0.053 | 0.837 |
| | | 0.4 | 21 | 0.066 | 0.866 | 0.053 | 0.827 |
| | | 0.5 | 22 | 0.064 | 0.873 | 0.051 | 0.834 |
| | | 0.6 | 22 | 0.064 | 0.851 | 0.051 | 0.816 |
| | | 0.7 | 22 | 0.064 | 0.850 | 0.051 | 0.820 |
| | (0.05,0.05) | 0.3 | 27 | 0.053 | 0.855 | 0.039 | 0.825 |
| | | 0.4 | 27 | 0.053 | 0.842 | 0.039 | 0.819 |
| | | 0.5 | 28 | 0.052 | 0.850 | 0.044 | 0.826 |
| | | 0.6 | 28 | 0.052 | 0.857 | 0.044 | 0.836 |
| | | 0.7 | 29 | 0.050 | 0.851 | 0.043 | 0.833 |
| DU (34, 56) | (0.03, 0.03) | 0.3 | 21 | 0.064 | 0.859 | 0.049 | 0.830 |
| | | 0.4 | 21 | 0.064 | 0.852 | 0.049 | 0.822 |
| | | 0.5 | 21 | 0.066 | 0.845 | 0.051 | 0.812 |
| | | 0.6 | 22 | 0.063 | 0.850 | 0.054 | 0.823 |
| | | 0.7 | 22 | 0.063 | 0.847 | 0.054 | 0.808 |
| | (0.05,0.05) | 0.3 | 27 | 0.063 | 0.843 | 0.055 | 0.808 |
| | | 0.4 | 27 | 0.063 | 0.830 | 0.055 | 0.808 |
| | | 0.5 | 28 | 0.064 | 0.858 | 0.058 | 0.834 |
| | | 0.6 | 28 | 0.064 | 0.841 | 0.058 | 0.819 |
| | | 0.7 | 29 | 0.063 | 0.853 | 0.054 | 0.828 |
| DU (10, 80) | (0.03, 0.03) | 0.3 | 22 | 0.068 | 0.852 | 0.054 | 0.818 |
| | | 0.4 | 23 | 0.061 | 0.860 | 0.041 | 0.819 |
| | | 0.5 | 23 | 0.066 | 0.847 | 0.046 | 0.810 |
| | | 0.6 | 24 | 0.068 | 0.851 | 0.049 | 0.814 |
| | | 0.7 | 24 | 0.064 | 0.842 | 0.046 | 0.807 |
| | (0.05,0.05) | 0.3 | 29 | 0.056 | 0.836 | 0.042 | 0.803 |
| | | 0.4 | 30 | 0.052 | 0.835 | 0.041 | 0.812 |
| | | 0.5 | 30 | 0.052 | 0.834 | 0.041 | 0.803 |
| | | 0.6 | 31 | 0.052 | 0.841 | 0.042 | 0.808 |
| | | 0.7 | 32 | 0.053 | 0.844 | 0.040 | 0.809 |

the required number of clusters \((N)\) are described as follows: we assume the average number of falls and the proportion of zeros to be \((1.21, 37.2\%)\) for the control group, as observed in the original study. It implies that \(p^*_1 = 12.1\%\) and \(\beta_1 = 0.19\). We assume that the intervention reduces the average number of falls to 1.01, which gives \(\beta_2 = -0.18\). We consider \(q = 0.3, 0.4, 0.5, 0.6,\) and 0.7 for sensitivity analysis, corresponding to \(p^*_2 = 20.0\%, 22.5\%, 24.9\%, 27.2\%,\) and 29.4\%, respectively. We further assume ICCs: \(\rho_s = \rho_u = 0.05\). Suppose the variability in cluster sizes is relatively small, say \(m_i \sim DU[127,147]\), the required number of clusters is 53 for \(q = 0.03\), 54 for \(q\) between 0.4 and 0.6, and 55 for \(q = 0.7\), respectively, based on the \(t\)-distribution approximation \((12)\). For a larger variability like the original study, say \(m_i \sim DU[37,237]\), the required number of clusters increases to 61 for \(q = 0.3\) and 0.4, and 62 otherwise.
TABLE 3 Comparison of $N^{(ZIP)}$ and $N^{(Poisson)}$ under randomly varying cluster sizes

|               | $(p_x, p_u)$ | $q$ | $N^{(ZIP)}$ | $\hat{\rho}^{(Poisson)}$ | $N^{(Poisson)}$ |
|---------------|-------------|-----|-------------|---------------------------|-----------------|
| DU (34, 56)   | (0.03, 0.03)| 0.3 | 18          | 0.022                     | 10              |
|               |             | 0.4 | 19          | 0.021                     | 10              |
|               |             | 0.5 | 19          | 0.023                     | 10              |
|               |             | 0.6 | 20          | 0.022                     | 10              |
|               |             | 0.7 | 20          | 0.021                     | 10              |
|               | (0.05, 0.05)| 0.3 | 24          | 0.037                     | 13              |
|               |             | 0.4 | 25          | 0.036                     | 13              |
|               |             | 0.5 | 25          | 0.036                     | 13              |
|               |             | 0.6 | 26          | 0.036                     | 13              |
|               |             | 0.7 | 27          | 0.038                     | 13              |
| DU (10, 80)   | (0.03, 0.03)| 0.3 | 20          | 0.022                     | 11              |
|               |             | 0.4 | 20          | 0.023                     | 11              |
|               |             | 0.5 | 21          | 0.022                     | 11              |
|               |             | 0.6 | 21          | 0.022                     | 11              |
|               |             | 0.7 | 22          | 0.023                     | 11              |
|               | (0.05, 0.05)| 0.3 | 27          | 0.035                     | 14              |
|               |             | 0.4 | 28          | 0.038                     | 15              |
|               |             | 0.5 | 28          | 0.035                     | 14              |
|               |             | 0.6 | 29          | 0.036                     | 14              |
|               |             | 0.7 | 30          | 0.036                     | 14              |

Note: Here $N^{(ZIP)}$ is the number of clusters calculated from formula (11), and $N^{(Poisson)}$ is calculated based on the Poisson distribution. $\hat{\rho}^{(Poisson)}$ is the estimated ICC for the Poisson distribution.

5 | CONCLUSION

In this study we present a sample size method for CRTs with zero-inflated count outcomes. It is developed based on GEE regression directly modeling the marginal mean of a ZIP outcome, which avoids the challenge of testing two intervention effects under traditional modeling approaches. We derive closed-form sample size formulas that properly account for zero inflation, ICCs due to clustering, unbalanced randomization, and variability in cluster size. We also introduce a new parameter $q$, which provides a straightforward decomposition of the overall intervention effect to facilitate communication with clinicians, as well as a natural framework to conduct sensitivity analysis. R code for sample size calculation and parameter estimations under GEE-type inference is included in the Supplementary Materials.

We evaluate the performance of the proposed sample size method through extensive simulation. The results show that the combination of calculating sample size based on the $t$-distribution ($N^{(t)}$) in experimental design, and implementing the “GEE-Jackknife” approach in data analysis, adequately controls the type I error and power at their nominal levels across all scenarios considered. We further show that traditional power analysis methods based on Poisson distribution tend to seriously underestimate sample sizes when the CRT has a zero-inflated count outcome.

The proposed method has some limitations. First, an independent working correlation was utilized to facilitate the derivation of a closed-form sample size solution. In situations with small or modest correlations, efficiency loss by using the independent working correlation structure is generally minimal. Under large correlations, however, this approach may suffer from efficiency loss and thus can be conservative. Future investigations may be conducted to explore whether a more sophisticated covariance structure can be used. Despite being conservative due to potential loss of efficiency, the proposed method enhances robustness against model mis-specification when information about the true correlation structure is lacking. Second, the ICCs required in the sample size formula are associated with $s_{ij}$ and $u_{ij}$, which are two
latent variables. Therefore, researchers may need to apply the moment-based estimators for ICC as described in Reference 34 using preliminary data that may be available for control group, leading to additional efforts in determining the appropriate values for ICCs of the latent variables in the design stage of a CRT.

The proposed sample size method is developed under the ZIP framework. In future research we will investigate its extension to other zero-inflated count models, such as the zero-inflated negative binomial model and the hurdle model.19

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID
Zhengyang Zhou https://orcid.org/0000-0002-8039-418X
Dateng Li https://orcid.org/0000-0002-4287-5337
Song Zhang https://orcid.org/0000-0002-9364-4053

REFERENCES
1. Murray DM, Pals SL, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. Am J Public Health. 2004;94:423-432.
2. Eldridge S, Kerry S. A Practical Guide to Cluster Randomised Trials in Health Services Research. Vol 120. Hoboken, NJ: John Wiley & Sons; 2012.
3. Murray DM. Design and Analysis of Group-Randomized Trials. New York: Oxford University Press; 1998.
4. Roig I, Perez S, Prieto G, et al. Cluster randomized trial in smoking cessation with intensive advice in diabetic patients in primary care. ITADI Study. BMC Public Health. 2010;10:58.
5. Fairall LR, Zwarenstein M, Bateman ED, et al. Effect of educational outreach to nurses on tuberculosis case detection and primary care of respiratory illness: pragmatic cluster randomised controlled trial. BMJ. 2005;331:750-754.
6. Martin CM, Doig GS, Heyland DK, Morrison T, Sibbald WJ. Multicentre, cluster-randomized clinical trial of algorithms for critical-care enteral and parenteral therapy (ACCEPT). CMAJ. 2010;179:197-204.
7. Cameron AC, Trivedi PK. Regression Analysis of Count Data. Vol 53. Cambridge, UK: Cambridge University Press; 2013.
8. Agresti A. Categorical Data Analysis. Vol 482. Hoboken, NJ: John Wiley & Sons; 2003.
9. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics. 1992;34:1-14.
10. Lewsey JD, Thomson WM. The utility of the zero-inflated Poisson and zero-inflated negative binomial models: a case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. Community Dent Oral Epidemiol. 2004;32:183-189.
11. Moghimbeigi A, Eshraghian MR, Mohammad K, Mcardle B. Multilevel zero-inflated negative binomial regression modeling for over-dispersed count data with extra zeros. J Appl Stat. 2008;35:1193-1202.
12. Kennedy C, Ioannidis G, Thabane L, et al. Successful knowledge translation intervention in long-term care: final results from the vitamin D and osteoporosis study (ViD OS) pilot cluster randomized controlled trial Trials. 2015;16:2141-11
13. Yang Z, Hardin JW, Addy CL. Testing overdispersion in the zero-inflated Poisson model. J Stat Plan Inf. 2009;139:3340-3353.
14. Amata A, Bhaumik D, Gibbons RD. Sample size determination for clustered count data. Stat Med. 2013;32:4162-4179.
15. Ahn C, Heo M, Zhang S. Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research. Boca Raton, FL: Chapman & Hall/CRC Press; 2014.
16. Liu J, Colditz GA. Relative efficiency of unequal versus equal cluster sizes in cluster randomized trials using generalized estimating equation models. Biometr J. 2018;60:616-638.
17. Wang J, Zhang S, Ahn C. Sample size calculation for count outcomes in cluster randomization trials with varying cluster sizes. Commun Stat Theory Methods. 2018;49(1):1-9.
18. Li D, Zhang S, Cao J. Sample size calculation for clinical trials with correlated count measurements based on the negative binomial distribution. Stat Med. 2019;38:5413-5427.
19. Mullaly J. Specification and testing of some modified count data models. J Econometr. 1986;33:341-365.
20. Preisser JS, Stamm JW, Long DL, Kincade ME. Review and recommendations for zero-inflated count regression modeling of dental caries indices in epidemiological studies. Caries Res. 2012;46:413-423.
21. Famoye F, Singh KP. Zero-inflated generalized Poisson regression model with an application to domestic violence data. J Data Sci. 2006;4:117-130.
22. Koning IM, Eijnden RJJM, Verdurmen JEE, Engels RCME, Vollebergh WAM. A cluster randomized trial on the effects of a parent and student intervention on alcohol use in adolescents four years after baseline; no evidence of catching-up behavior. Addict Behav. 2013;38:2032-2039.
23. Williamson JM, Lin H, Lyles RH, Hightower AW. Power calculations for ZIP and ZINB models. J Data Sci. 2007;5:519-534.
24. Channouf N, Fredette M, MacGibbon B. Power and sample size calculations for Poisson and zero-inflated Poisson regression models. Comput Stat Data Anal. 2014;72:241-251.
SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhou Z, Li D, Zhang S. Sample size calculation for cluster randomized trials with zero-inflated count outcomes. Statistics in Medicine. 2022;41(12):2191-2204. doi: 10.1002/sim.9350

APPENDIX A. DERIVATION OF $V_1$ AND $V_2$

$V_1$ and $V_2$ are derived in the similar manner. Here we present the derivation of $V_1$. First rewrite $V_1$ as

$$V_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \times \frac{1}{\left(1 + \frac{\mu_1^*}{1 - \mu_1^*} \right)^2} \sum_{m \in M} g(m) E \left[ \sum_{j=1}^{m} \sum_{j' = 1}^{m} (y_{ij} - \mu_1^*)(y_{ij'} - \mu_1^*) \right].$$

For $E \left[ \sum_{j=1}^{m} \sum_{j' = 1}^{m} (y_{ij} - \mu_1^*)(y_{ij'} - \mu_1^*) \right]$, we have

$$E \left[ \sum_{j=1}^{m} \sum_{j' = 1}^{m} (y_{ij} - \mu_1^*)(y_{ij'} - \mu_1^*) \right] = \sum_{j=1}^{m} E(y_{ij} - \mu_1^*)^2 + 2 \sum_{j=1}^{m-1} \sum_{j' = j+1}^{m} E(y_{ij} - \mu_1^*)(y_{ij'} - \mu_1^*).$$
It is clear that
\[
\sum_{j=1}^{m} E(y_{ij} - \mu_1^*)^2 = \sum_{j=1}^{m} \text{Var}(y_{ij}) = m \left[ \mu_1^* + \frac{p_1^*}{1 - p_1^*} \mu_1^* \right].
\]

On the other hand, we have
\[
2 \sum_{j=1}^{m-1} \sum_{j' = j+1}^{m} E \left[ (y_{ij} - \mu_1^*)(y_{ij'} - \mu_1^*) \right] = 2 \sum_{j=1}^{m-1} \sum_{j' = j+1}^{m} \text{Cov}(y_{ij}, y_{ij'}) = 2(m^2 - m)\text{Cov}(y_{ij}, y_{ij'}). \]

If \(s_{ij} = s_{ij'} = 1\),
\[
E \left[ (y_{ij} - \mu_1^*)(y_{ij'} - \mu_1^*) \right] = E \left[ (0 - \mu_1^*)(0 - \mu_1^*) \right] = \mu_1^{*2};
\]
if \(s_{ij} = 1, s_{ij'} = 0\),
\[
E \left[ (y_{ij} - \mu_1^*)(y_{ij'} - \mu_1^*) \right] = E \left[ (0 - \mu_1^*)(u_{ij'} - \mu_1^*) \right]
= -\mu_1^* E \left[ u_{ij'} - \lambda_1^* + p_1^* \lambda_1^* \right]
= -\mu_1^{*2} \frac{p_1^*}{1 - p_1^*},
\]
where \(\lambda_1^* = \frac{\mu_1^*}{1 - p_1^*}\);
if \(s_{ij} = s_{ij'} = 0\),
\[
E \left[ (y_{ij} - \mu_1^*)(y_{ij'} - \mu_1^*) \right] = E \left[ (u_{ij} - \mu_1^*)(u_{ij'} - \mu_1^*) \right]
= E \left[ (u_{ij} - \lambda_1^* + p_1^* \lambda_1^*)(u_{ij'} - \lambda_1^* + p_1^* \lambda_1^*) \right]
= \rho_u \lambda_1^* + p_1^{*2} \lambda_1^{*2}
= \frac{\rho_u \mu_1^*}{1 - p_1^*} + \frac{p_1^{*2} \mu_1^{*2}}{(1 - p_1^*)^2}.
\]

It is easy to verify that \(\text{Prob}(s_{ij} = s_{ij'} = 1) = p_1^{*2} + p_1^*(1 - p_1^*)\rho_u\); \(\text{Prob}(s_{ij} = 0, s_{ij'} = 1) = (1 - p_1^*)p_1^*(1 - \rho_u)\); \(\text{Prob}(s_{ij} = s_{ij'} = 0) = (1 - p_1^*)(1 - p_1^* + \rho_u p_1^*)\).

In addition, it is clear that \(\sum_{m \in \mathcal{A}} g(m)m = \eta_m\) and \(\sum_{m \in \mathcal{A}} g(m)(m^2 - m) = \eta^2_m + \sigma^2_m - \eta_m\). Putting together, we have
\[
V_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \left[ \begin{array}{c} \mu_1^* + \frac{p_1^*}{1 - p_1^*} \mu_1^{*2} \\ \eta_m \left( \frac{p_1^*}{1 - p_1^*} \mu_1^* \right)^2 + (\eta^2_m + \sigma^2_m - \eta_m) \left( \frac{\zeta_1}{1 - p_1^*} \mu_1^{*2} \right)^2 \end{array} \right],
\]
with
\[
\zeta_1 = \mu_1^{*2} p_1^*(1 + p_1^*)\rho_u + \mu_1^* p_1^*(1 - p_1^* + p_1^* \rho_u) + \frac{\mu_1^{*2} p_1^* \rho_u}{1 - p_1^*}.
\]

Hence we complete the derivation of \(V_1\).