Power Calculations for ZIP and ZINB Models

John M. Williamson[1], Hung-Mo Lin[2], Robert H. Lyles[3] and
Allen W. Hightower[1]
[1]*Centers for Disease Control and Prevention,*
[2]*Penn State College of Medicine and* [3]*Emory University*

Abstract: We present power calculations for zero-inflated Poisson (ZIP) and
zero-inflated negative-binomial (ZINB) models. We detail direct computa-
tions for a ZIP model based on a two-sample Wald test using the expected
information matrix. We also demonstrate how Lyles, Lin, and Williamson's
method (2006) of power approximation for categorical and count outcomes
can be extended to both zero-inflated models. This method can be used
for power calculations based on the Wald test (via the observed information
matrix) and the likelihood ratio test, and can accommodate both categori-
cal and continuous covariates. All the power calculations can be conducted
when covariates are used in the modeling of both the count data and the
"excess zero" data, or in either part separately. We present simulations to
detail the performance of the power calculations. Analysis of a malaria study
is used for illustration.

Key words: Likelihood ratio test, non-central chi-square distribution, power,
sample size, Wald test, zero-inflated negative binomial model, zero-inflated
Poisson model.

1. Introduction

Often the outcome of interest in public health and medical studies is a count
variable, which is usually assumed to follow the Poisson distribution and can
be modeled accordingly. However, the count variable may contain excess zeroes
above what is to be expected from the Poisson model. These excess zeroes may
be due to 1) the presence of a subpopulation with only zero counts, 2) overdis-
persion, or 3) chance (Campbell, Machin, and D'Arcangues, 1991). A common
approach for analyzing such data is the zero-inflated Poisson (ZIP) model (Mul-
lahy, 1986; Lambert, 1992), which is an extension of Cohen's (1960) modified
Poisson distribution. ZIP modeling has been used to analyze data on caries pre-
vention (Bohning, Dietz, Schlattmann, Mendonca, and Kirchner, 1999), early
growth failure in children (Cheung, 2002), and sudden infant death syndrome

(Dalrymple, Hudson, and Ford, 2003), as well as many other diseases. ZIP modeling also can accommodate the extent of individual exposure (Lee, Wang, and Yau, 2001). A score test has been proposed for determining whether the number of zeros is too large for a Poisson distribution to fit the data well, indicating that a ZIP model may be a better choice (van den Broek, 1995). ZIP models also have been extended to the bivariate (Walhin, 2001) and multivariate settings (Li, Lu, Park, Brinkley, and Peterson, 1999). Hall (2000) and Yau and Lee (2001) have used random-effects approaches to extend the ZIP model to analyze longitudinal count data.

If count data are overdispersed such that the variance of the count variable is greater than the mean, then the Poisson assumption is violated. A negative binomial distribution may then be used for modeling purposes, as it uses an additional parameter in describing the variance of the count variable. If the data are still zero inflated, a zero-inflated negative binomial (ZINB) model may be fit. A score test for a ZIP regression model against a ZINB alternative (Ridout, Hinde, and Demetrio, 2001) has been proposed. Hall and Berenhaut (2002) provide an alternative to Ridout, Hinde, and Demetrio's (2001) score test by testing that the variance and covariance components associated with random effects in a ZIP model are equal to zero, implying that a ZINB model does not fit the data better than the ZIP alternative.

Studies without a sufficient sample size often will result in a failure to detect a significant effect when it exists; however, there is often a high cost associated with recruiting and evaluating large samples of subjects. This consideration makes sample size (power) calculations a crucial step in designing clinical research and public health studies, many of which are known to have a ZIP- or ZINB-distributed response. Sample size calculations are especially important for zero-inflated models because a larger sample size is required to detect a significant effect with these models than with the standard Poisson or negative-binomial models. To date, methods for power calculations for ZIP and ZINB models are scarce.

Here we present an example of a study with a zero-inflated count response. Weekly household mosquito counts were obtained for a longitudinal study of malaria risk factors in Western Kenya (Bloland et al., 1999). In this area, some houses are reputed to be "malaria houses" - where the risk of disease is great. A goal of the study was to see if basic sanitation and construction practices would affect malaria risk. A major outcome of the study was the weekly household mosquito count, which has an excess number of 0 values as some houses never have mosquitos.

In Section 2, we review the ZIP and ZINB models. We use Lyles, Lin, and Williamson's (2006) approach for computing power for generalized linear models

to compute power for both zero-inflated models in Section 3. The power calculations can be based on the Wald or the likelihood ratio tests. We also present power computations for a two-group comparison with a ZIP model based on a Wald test where the variance-covariance matrix is calculated from the expected information matrix. The latter approach was not feasible for the ZINB model because of the difficulty in taking the expected value of the second derivative matrix. We present simulation studies to detail the methods' performance in Section 4 and illustrate their use with a public health example in Section 5. We conclude with a short discussion on the merits of these sample-size calculations.

2. Zero-Inflated Models

Let the response Y_i denote a non-negative integer count for the ith observation, $i = 1, \cdots, N$. The probability of an excess zero is denoted by π_i, $0 \leq \pi_i \leq 1$. Following Cheung (2002), the random variable Y_i follows a ZIP distribution if

$$Pr(Y_i = y_i) = \begin{cases} \pi_i + (1 - \pi_i)e^{-\lambda_i} & \text{if } y_i = 0 \\ (1 - \pi_i)\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} & \text{if } y_i > 0 \end{cases} \tag{2.1}$$

for $i = 1, \cdots, N$. The mean and variance of the ZIP random variable are $E(Y_i) = (1 - \pi_i)\lambda_i$ and $Var(Y_i) = (1 - \pi_i)\lambda_i(1 + \pi_i\lambda_i)$. If Y_i follows a ZINB distribution, then

$$Pr(Y_i = y_i) = \begin{cases} \pi_i + (1 - \pi_i)\left(\frac{1}{1+\kappa\lambda_i}\right)^{\kappa^{-1}} & \text{if } y_i = 0 \\ (1 - \pi_i)\frac{\Gamma(\kappa^{-1} + y_i)}{\Gamma(\kappa^{-1})(y_i!)}\left(\frac{\kappa\lambda_i}{1+\kappa\lambda_i}\right)^{y_i}\left(\frac{1}{1+\kappa\lambda_i}\right)^{\kappa^{-1}} & \text{if } y_i > 0 \end{cases} \tag{2.2}$$

for $i = 1, \cdots, N$. The mean and variance of the ZINB random variable are $E(Y_i) = (1 - \pi_i)\lambda_i$ and $Var(Y_i) = (1 - \pi_i)\lambda_i(1 + (\kappa + \pi_i)\lambda_i)$, where κ is an overdispersion parameter. The ZINB model reduces to the ZIP model as $\kappa \to 0$ (that is, equation 2 and equation 1 are then equivalent).

For both models, we assume that π_i will be modeled with logistic regression, $logit(\pi_i) = \boldsymbol{X}_i\boldsymbol{\beta}$, where $\boldsymbol{X}_i$ is a $1 \times p$ row vector of covariates (including an intercept) and $\boldsymbol{\beta}$ is the corresponding $p \times 1$ column vector of parameters. This is a natural choice for a binary response and results in an odds ratio interpretation for the parameters. We model the mean of the count part as $log(\lambda_i) = \boldsymbol{Z}_i\boldsymbol{\gamma}$, where $\boldsymbol{Z}_i$ is a $1 \times q$ row vector of covariates and $\boldsymbol{\gamma}$ is the corresponding $q \times 1$ column vector of parameters. Let $\boldsymbol{\theta} = [\, \boldsymbol{\beta}', \boldsymbol{\gamma}'\,]'$. The covariate vectors $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ may contain the same or differing predictors.

3. Power Calculations for Zero-Inflated Models

Lyles, Lin, and Williamson (2006) proposed a simple and flexible method for estimating conditional power (that is, power given any prespecified fixed covariate design matrix) for binary, ordinal, and count outcomes. This method requires only standard software for fitting the desired generalized linear model. The model is fit to a representative expanded data set using easily calculated weights that represent response probabilities given the assumed values of the parameters. The resulting variance-covariance matrix is used for power calculations based on the Wald test. Power based on the likelihood ratio statistic can be approximated by refitting the model under the null hypothesis.

Lyles, Lin and Williamson's (2006) approach proceeds as follows. Assume that the zero-inflated response variable Y takes J possible values $0, 1, \cdots, J-1$ where J is chosen such that the $Pr(Y \geq J | X_i = x_i, Z_i = z_i)$ for any specific vectors x_i and z_i in their respective design matrices is negligible. For example, one could use the criterion that $Pr(Y < J | X_i = x_i, Z_i = z_i) \geq 0.999$ for each distinct covariate set. The representative expanded data set proposed here represents a generalization of an "exemplary" data set proposed for other power calculations (O'Brien, 1986; O'Brien, 1988; and O'Brien and Muller, 1993). The expanded data set is comprised of one record for each of the J response values per combination of covariate values with an associated weight $w_{ij} = Pr(Y_i = j | X_i = x_i, Z_i = z_i)$ easily calculated from (1) and (2). We are interested in testing the following hypothesis:

$$H_0 : \mathbf{H}\boldsymbol{\theta} = \mathbf{h_0} \quad \text{versus} \quad H_A : \mathbf{H}\boldsymbol{\theta} \neq \mathbf{h_0} \tag{3.1}$$

where $\mathbf{H}$ is an $(h \times (p+q))$ matrix of full rank and $\mathbf{h_0}$ is a $(h \times 1)$ vector of constant terms.

The Wald and likelihood ratio (LR) test statistics are given by

$$T_W = (\mathbf{H}\widehat{\boldsymbol{\theta}} - \mathbf{h}_0)'[\mathbf{H}\widehat{\mathbf{var}}(\widehat{\boldsymbol{\theta}})\mathbf{H}']^{-1}(\mathbf{H}\widehat{\boldsymbol{\theta}} - \mathbf{h}_0)$$

and

$$T_{LR} = -2[l(\widehat{\boldsymbol{\theta}}^*) - l(\widehat{\boldsymbol{\theta}})],$$

where $\widehat{\boldsymbol{\theta}}$ is the unrestricted maximum likelihood estimate of $\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^*$ is the corresponding maximum likelihood estimator under H_0, and $l(.)$ denotes the log-likelihood function. Under H_0, both test statistics are asymptotically distributed as central chi-square random variables with h degrees of freedom. Under H_A, and following Wald (1943) and Stroud (1973), T_W is asymptotically distributed as non-central $\chi^2_{h,\lambda}$, where the non-centrality parameter λ is $(\mathbf{H}\boldsymbol{\theta} - \mathbf{h}_0)'[\mathbf{H}\mathbf{var}(\widehat{\boldsymbol{\theta}})\mathbf{H}']^{-1}(\mathbf{H}\boldsymbol{\theta} - \mathbf{h}_0)$. For the likelihood ratio test, T_{LR} is asymptotically distributed

as non-central $\chi^2_{h,\nu}$, where the non-centrality parameter is $\nu = -2[l^*(\boldsymbol{\theta}) - l(\boldsymbol{\theta})]$, $l(\boldsymbol{\theta})$ being the log-likelihood under the true $\boldsymbol{\theta}$, and $l^*(\boldsymbol{\theta})$ is the loglikelihood when $\boldsymbol{\theta}$ is evaluated under H_0 (Wald, 1943; Stuart, Ord, and Arnold, 1999). Let α represent the specified type-I error rate and $\chi^2_{h:1-\alpha}$ the critical value from the central χ^2_h distribution. The power for testing H_0 with the Wald or likelihood test is

$$1 - \eta = \int_{\chi^2_{h:1-\alpha}}^{\infty} f(u; h, c)\, du$$

where η is the type-II error rate, $f(u; h, c)$ is the probability density function of the non-central $\chi^2_{h,c}$ distribution, and c is the appropriate degrees of freedom for either test.

The expanded data set is created as follows. For $i = 1, \cdots, N$, specify $\mathbf{x}_i$ and $\mathbf{z}_i$ and calculate the weight w_{ij} for specified $\boldsymbol{\theta}$ (and also κ for the ZINB model). Choose J such that the probability that Y is greater than or equal to J is negligible. Each record in the data set will correspond to an observation with response j $(j = 0, 1, \cdots, J - 1)$. For example, assume that we are calculating the power for a sample of $N = 100$ and have chosen $J = 16$. Additionally, we assume that one covariate is used in the "excess zero" modeling, x_i, and a different covariate is used in the count modeling, z_i. The expanded data set will include $N \times J = 1600$ records with the 16 records for the ith experimental unit given as follows:

X_i	Z_i	Y_{ij}	W_{ij}
x_i	z_i	0	w_{i0}
x_i	z_i	1	w_{i1}
x_i	z_i	2	w_{i2}
		$\cdots$	
x_i	z_i	14	w_{i14}
x_i	z_i	15	w_{i15}

The third column represents the response and $\sum_{j=0}^{15} w_{ij} \approx 1.0$. If hypothetically x_i and z_i were specified to be 2 different binary covariates (50% in each category for each covariate), then the first 25 experimental units (400 records) would have $x_i = z_i = 0$, the next 25 $x_i = 0$ and $z_i = 1$, the following 25 $x_i = 1$ and $z_i = 0$, and the last 25 $x_i = z_i = 1$. The expanded data set can be shortened for categorical covariates by multiplying the appropriate weights by the number of observations with the corresponding covariate pattern. For the above example, the shortened data set would have 64 (16×4 distinct covariate patterns) records with the weights multiplied by 25.

When the predictor of interest is a continuous variable, we use a representative covariate data set along with the expanded data set. Let x_i be a realization of a continuous covariate X whose cumulative distribution function $F(x)$ is assumed known. Using the Blom adjustment (Blom, 1958), we create a representative covariate data set consisting of the expected quantiles: $x_i = F^{-1}\left(\frac{i-0.375}{N+0.25}\right)$. An example is found in the second set of simulations in Section 4 where such a data set is illustrated in a table. See Lyles, Lin, and Williamson (2006) for a more detailed explanation.

3.1 Power calculations for 2-sample test with ZIP data

Here we demonstrate that power for ZIP models can also be calculated using the traditional approach where the variance-covariance matrix of parameter estimates is calculated from the expected information matrix. In contrast, a similar approach for the ZINB model is not feasible due to the difficulty in taking the expected value of its second derivative matrix. Although this method is extendible to multiple groups, we focus here on a two-group comparison and present power calculations for ZIP models using the Wald test. For most sample size calculations conducted for medical and public health studies, the main hypothesis of interest is usually a comparison between two groups, such as treatment (new drug versus standard drug or placebo) or gender (male versus female). A binary covariate x_i will be used in the modeling of both parts of the ZIP model. The probability π_i of an excess zero will be modeled as $logit(\pi_i) = \beta_0 + \beta_1 x_i$ and the mean of the Poisson part as $log(\lambda_i) = \gamma_0 + \gamma_1 x_i$. Therefore, the parameters of interest here are $\boldsymbol{\theta} = [\beta_0, \beta_1, \gamma_0, \gamma_1]'$.

We present the score functions and the information matrix for this ZIP model in Appendix 1. Let $\widehat{\boldsymbol{\theta}}$ be the solution to the score equations. Then, $\mathbf{var}(\widehat{\boldsymbol{\theta}}) = (1/N)\mathbf{I}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta})$, which can be estimated by substituting $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$. Specify the hypothesis for $\boldsymbol{\theta}$ as (3). For example, if the hypothesis of interest is that the two groups ($x_i = 0, 1$) have the same probability of an excess zero, then $\mathbf{H} = [\begin{array}{cccc} 0 & 1 & 0 & 0 \end{array}]$ and $h_0 = 0$. If one is interested in testing for an overall difference between the two groups, then one may use

$$\mathbf{H} = \left[\begin{array}{cccc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

and $\mathbf{h_0} = [\begin{array}{cc} 0 & 0 \end{array}]$. One only has to specify the sample sizes (N_0 and N_1), the probabilities of an excess zero (π_0 and π_1), and the means of the Poisson counts (λ_0 and λ_1) for the two groups in order to conduct these power calculations, as $\boldsymbol{\theta}$ is easily solved in terms of these parameters for the two-group comparison. The advantage here is that it is easier for a medical investigator to specify probabilities

Table 1: Power (%) to test $H_0 : [\beta_1, \gamma_1]' = \mathbf{0}$ for a two-sample ZIP model at $\alpha = 0.05$ (two sided) with $N_0 = N_1 = 100$.

$$\pi_0 = 0.15$$

	$\pi_1 = 0.20$				$\pi_1 = 0.25$				$\pi_1 = 0.25$		
		λ_1				λ_1				λ_1	
λ_0	4.0	5.0	6.0	λ_0	4.0	4.5	5.0	λ_0	10.0	11.0	12.0
4.0	11.3	77.6	99.9	4.0	30.0	47.4	83.8	10.0	32.8	64.7	97.2
	11.4	78.1	99.9		31.0	48.6	84.7		33.6	65.2	97.2
	12.1	80.6	99.9		28.4	47.0	88.2		30.2	73.7	99.5
4.5	34.6	30.0	97.1	4.5	53.5	31.1	47.8	10.5	41.9	41.5	85.5
	34.8	30.1	97.3		54.4	32.0	48.8		42.7	42.3	85.7
	33.5	28.1	98.6		52.0	31.0	51.6		40.7	40.7	88.3
5.0	80.4	11.8	71.8	5.0	87.5	51.6	31.7	11.0	64.5	32.8	62.4
	80.8	11.8	72.1		88.1	52.4	32.6		65.2	33.6	62.9
	75.3	10.9	73.9		83.5	48.4	28.8		59.4	27.1	69.1

$$\pi_0 = 0.45$$

	$\pi_1 = 0.50$				$\pi_1 = 0.55$				$\pi_1 = 0.60$		
		λ_1				λ_1				λ_1	
λ_0	5.5	6.0	6.5	λ_0	11.0	12.0	13.0	λ_0	5.0	5.5	6.0
4.0	88.5	98.6	99.9	10.0	44.8	85.4	99.0	4.0	76.8	93.7	99.1
	89.0	98.8	99.9		44.8	85.4	99.0		77.1	93.9	99.2
	88.7	98.7	99.9		44.7	85.5	99.1		78.2	93.4	98.3
4.5	54.2	86.0	97.9	10.5	28.0	65.3	94.7	4.5	54.1	75.7	92.6
	54.4	86.4	98.1		28.1	65.3	94.7		54.5	75.9	92.7
	53.1	85.1	97.5		26.3	71.2	98.7		54.2	74.0	92.1
5.0	19.5	51.2	83.3	11.0	22.5	42.9	82.7	5.0	45.3	53.9	74.3
	19.5	51.3	83.6		22.6	43.0	82.6		45.8	54.3	74.6
	20.5	52.9	84.2		20.9	42.3	83.3		48.0	55.9	74.7

and mean counts then parameters on the logit or log scales.

4. Simulations

To assess the performance of our proposed power calculations for both approaches, we conducted several studies using simulated data. For the first set of simulations, we randomly generated ZIP-distributed data for 200 observations with a binary covariate ($x = 0, 1$ for 100 observations each) for each simulated data set. The parameters used to generate the data ($\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$) were specified

Table 1 (continued): Power (%) to test $H_0 : [\beta_1, \gamma_1]' = \mathbf{0}$ for a two-sample ZIP model at $\alpha = 0.05$ (two sided) with $N_0 = N_1 = 100$.

$$\pi_0 = 0.75$$

$\pi_1 = 0.80$				$\pi_1 = 0.85$				$\pi_1 = 0.90$		
	λ_1				λ_1				λ_1	
λ_0	5.5	6.0	6.5	λ_0	11.0	12.5	14.0	λ_0	5.0	5.5	6.0
4.0	55.7	77.6	91.4	10.0	41.3	74.0	95.7	4.0	75.3	83.4	90.7
	56.0	78.3	92.0		42.0	74.1	95.5		78.1	85.2	91.5
	56.2	78.7	91.3		39.3	74.2	95.3		78.3	86.9	92.6
4.5	30.4	52.8	74.7	10.5	34.9	61.3	90.2	4.5	69.2	75.1	82.8
	30.4	53.0	75.1		35.7	61.6	89.9		72.9	78.1	84.7
	30.8	50.1	73.3		33.1	60.2	89.4		70.0	75.8	86.5
5.0	15.2	29.0	50.0	11.0	32.8	49.4	81.5	5.0	67.4	69.3	74.8
	15.2	29.0	50.1		33.6	50.0	81.3		71.4	73.1	77.9
	16.2	28.8	48.6		33.1	48.5	80.4		69.3	71.7	78.4

The calculated power using the Wald test with expected information matrix is the main entry (bold). The middle entry is the calculated power using Lyles et al.'s method based on the likelihood ratio test (italicized). The empirical power based on the Wald test from the 1000 simulations is presented below.

in terms of $\pi_0, \pi_1, \lambda_0,$ and λ_1. One thousand data sets were generated for various (but not exhaustive) combinations of $\pi_0, \pi_1, \lambda_0,$ and λ_1. See Table 1. Power calculations were conducted using Lyles, Lin, and Williamson's (2006) approach (both the Wald and likelihood ratio tests) and the proposed two-sample power computation. The data sets were analyzed with the ZIP model that was used to generate the data (i.e., the binary covariate was incorporated into both the logit and Poisson parts of the model), so that the model was correctly specified. To calculate the empirical power for each set of simulations, we estimated $\boldsymbol{\beta} = [\beta_0, \beta_1]'$ and $\boldsymbol{\gamma} = [\gamma_0, \gamma_1]'$ to test $H_0 : \beta_1 = \gamma_1 = 0$. The resulting Wald test statistic was compared to a chi-square distribution with two degrees of freedom. Then the empirical power based on the Wald test was calculated as the number of data sets resulting in the rejection of $H_0 : \beta_1 = \gamma_1 = 0 \, (\alpha = 0.05)$ divided by 1000.

Table 1 summarizes the simulation results. Of the 81 scenarios, the calculated power for the two-sample Wald test based on the expected information matrix was less than the empirical power 40 times, greater than the empirical power 39 times, and equal to the empirical power twice. The maximum difference between the calculated power and the empirical power was 0.090 ($\pi_0 = 0.15, \pi_1 = 0.25, \lambda_0 = 10.0, \lambda_1 = 11.0$), but there were only six scenarios (out of 81) for which the calculated power was different from the empirical power by more than 0.05.

Table 2: Expanded data set of 3100 lines $(N \times J)$ with $N = 100$ and $J = 31$ for ZINB simulation with $x_i \sim N(0,1)$. The excess zero part of the model is $logit(\pi_i) = -0.406 + 0.65x_i$ and the count part is $log(\lambda_i) = 1.609 + 0.25x_i$.

Obs	i	x_i	y_j	weight(w_{ij})	Obs	i	x_i	y_j	weight(w_{ij})
1	1	-2.4986	0	0.2197	3039	99	2.1392	0	0.7300
2	1	-2.4986	1	0.1806	3040	99	2.1392	1	0.0059
		.					.		
		.					.		
30	1	-2.4986	29	0.0000	3068	99	2.1392	29	0.0001
31	1	-2.4986	30	0.0000	3069	99	2.1392	30	0.0001
32	2	-2.1392	0	0.2279	3070	100	2.4986	0	0.7730
33	2	-2.1392	1	0.1580	3071	100	2.4986	1	0.0038
		.					.		
		.					.		
61	2	-2.1392	29	0.0000	3099	100	2.4986	29	0.0002
62	2	-2.1392	30	0.0000	3100	100	2.4986	30	0.0001

The results were excellent for moderate and large values of π. The largest discrepancies between the calculated power using the Wald test with the expected information matrix and the empirical power were for small values of π.

Power calculations using Lyles, Lin, and Williamson's (2006) method with the Wald test (observed information matrix) were not shown because the difference between this approach and the power calculation using the Wald test with the variance-covariance matrix calculated from the expected information matrix never exceeded 0.2%. Power calculations using Lyles, Lin, and Williamson's method with the likelihood ratio test were very similar to those based on the Wald test. In the instances where there was a slight difference, the power calculations based on the likelihood ratio test tended to be higher than those based on the Wald test.

For a second set of simulations, we randomly generated ZINB-distributed data (Krishnamoorthy, 2001) with samples of size 100 and 500. The probability π_i of an excess zero was modeled as $logit(\pi_i) = \beta_0 + \beta_1 x_i$ and the mean of the count part as $log(\lambda_i) = \gamma_0 + \gamma_1 x_i$. The covariate x_i was distributed as a $N(0,1)$ random variable or a $U(-0.5, 0.5)$ random variable. The parameters used to generate the data were $\beta_0 = -0.4055$ (corresponding to 40% excess zeros when $x_i = 0$), and $\gamma_0 = 1.6094$ (corresponding to a mean count of 5 when $x_i = 0$). The parameters β_1 and γ_1 varied according to the sample size and distribution of x_i. In general, we chose β_1 and γ_1 such that the resulting power is within practical range. Two thousand data sets were generated for each combination of the parameters. We found that taking $J = 31$ for both the $N(0,1)$ and $U(-0.5, 0.5)$ distributions ensured that the w_{ij}'s summed very close to one for each x_i. We used the criterion

that $\sum_{j=0}^{J-1} w_{ij} \geq 0.999 \,\forall\, x_i$. For specified $\boldsymbol{\theta}$, the criterion can be checked directly for each unique x_i. Alternatively, it is often easy to identify which x_i assigns highest probability to large Y, in which case one could choose J to ensure that the criterion holds for that x_i. See Table 2 for a partial listing of the expanded data set of size $100 \times 31 = 3100$ for the power calculation when $N = 100$ and x_i is distributed as a $N(0, 1)$ random variable using the Blom adjustment.

Table 3: Power (%) to test H_0 under a ZINB model with a continuous covariate at $alpha = 0.05$ (two sided). The models for the excess zero and count parts are $\mathrm{logit}(\pi_i) = -0.4055 + \beta_1 x_i$ and $\log(\lambda_i) = 1.6094 + \gamma_1 x_i$, respectively, with $\kappa = 0.2$.

		$N = 100$		$N = 500$	
Distribution of x_i		$N(0, 1)$	$U(-0.5, 0.5)$	$N(0, 1)$	$U(-0.5, 0.5)$
True		$\beta_1 = 0.65$	$\beta_1 = 2.0$	$\beta_1 = 0.25$	$\beta_1 = 0.9$
Parameters		$\gamma_1 = 0.25$	$\gamma_1 = 0.85$	$\gamma_1 = 0.1$	$\gamma_1 = 0.45$
Hypothesis H_0	Test				
$\beta_1 = \gamma_1 = 0$	Wald	0.885	0.881	0.883	0.966
		(0.907)	(0.911)	(0.886)	(0.971)
	LRT	0.915	0.900	0.890	0.967
		(0.924)	(0.921)	(0.889)	(0.972)
$\beta_1 = 0$	Wald	0.712	0.674	0.718	0.758
		(0.764)	(0.694)	(0.753)	(0.763)
	LRT	0.792	0.730	0.732	0.769
		(0.802)	(0.726)	(0.755)	(0.769)
$\gamma_1 = 0$	Wald	0.765	0.782	0.732	0.921
		(0.756)	(0.790)	(0.719)	(0.924)
	LRT	0.743	0.759	0.729	0.917
		(0.746)	(0.784)	(0.719)	(0.923)

The calculated power using the Wald and Likelihood Ratio tests are the main entries. The empirical powers from the 2000 simulations are presented below each entry in parentheses.

See Table 3 for the simulation results. Power calculations were conducted using Lyles, Lin, and Williamson's (2006) approach (both the Wald and likelihood ratio tests) for the tests $H_0 : \beta_1 = \gamma_1 = 0$, $H_0 : \beta_1 = 0$ and $H_0 : \gamma_1 = 0$. The empirical power was calculated as the number of data sets resulting in the rejection of the appropriate null hypothesis ($\alpha = 0.05$) divided by 2000, for the respective Wald or likelihood ratio test. Both sets of simulations were conducted via SAS IML[1]. Overall, the results were satisfactory. The largest difference be-

[1]SAS Institute, Inc. (1995). SAS/IML Software: Changes and Enhancements through Re-

tween calculated power and empirical power for data sets of size 100 was 0.052 for the Wald test ($x_i \sim N(0,1)$ and testing $H_0 : \beta_1 = 0$). For data sets of size 100 and the likelihood ratio test, there were no instances where the calculated power was different than the empirical power by more than 0.03. There were no instances where the calculated power was different than the empirical power by more than 0.04 for either test and data sets of size 500. For sample size of 100, the likelihood ratio test had greater power than the Wald test for $H_0 : \beta_1 = 0$, and the opposite was true for $H_0 : \gamma_1 = 0$. However, these differences were negligible when the sample size increased to 500.

Table 4: Mosquito count from 492 houses in Western Kenya

| | Separate Pit Latrine | |
Count	Yes	No
0	153	133
1	39	31
2	24	18
3	11	17
4	5	14
5	11	3
6	6	4
7	2	1
8	2	1
9	2	0
10	2	0
11	2	0
12	1	0
13	2	1
14	3	0
18	0	1
19	0	1
25	0	1
27	1	0
Total	266	226

lease 6.11. SAS Institute, Inc.: Cary, NC.

5. Western Kenya Malaria Cohort Study

Suppose that we want to calculate the sample size for a new study based on the parameter values estimated from the data collected in a similar previous study. One of the objectives of the Western Kenya Malaria Cohort study was to determine if basic household construction and sanitation practices were related to malaria risk (Bloland et al., 1999). All participating households were visited and information on construction, animal ownership, and sanitation practices (such as the presence of a separate pit latrine structure) were recorded. Since only two of the many species of mosquitoes in this area transmit malaria, each participating household then had weekly mosquito trapping sessions, using a standard CDC-developed mosquito trap during the time that these mosquitoes are actively feeding (generally 11:00 PM to 2:00 AM). Trapped mosquitoes were separated by species and malaria-transmitting species were counted to give the dependent measure used in this analysis. Before the initiation of the study, local residents had commented that certain homes were "malaria houses" while other nearby homes "had no mosquitoes" for reasons that were not well known. A simple change of sanitation or home construction procedure that resulted in a decrease to the inhabitants' likelihood of getting malaria would be of tremendous benefit. Virtually all infants in this area get malaria multiple times before their first birthday. Of particular interest was whether or not a separate pit latrine structure, which has clear sanitation benefits, was associated with any difference in the numbers of mosquitoes that transmit malaria.

The mosquito counts are from the first weekly visit to each house in September, 1994 and are presented in Table 4. In the sample of 492 houses, 286 (58%) had no mosquitoes, much greater than the expected count of 111 predicted from a standard Poisson distribution. Here we are concerned with whether mosquito counts are the same for houses with a separate pit latrine as those without one. We fit a ZIP model with the covariate $x = 0, 1$ corresponding to whether or not the house had a separate pit latrine. The resulting parameter estimates were $\widehat{\boldsymbol{\beta}} = [0.279, -0.020]'$ and $\widehat{\boldsymbol{\gamma}} = [1.136, 0.171]'$, which yielded $\widehat{\pi} = [0.569, 0.564]'$ and $\widehat{\lambda} = [3.113, 3.693]'$. The joint test of $H_0 : [\beta_1, \gamma_1]' = [0, 0]'$ was nonsignificant ($\chi^2_2 = 4.63$, p-value= 0.099) as was the marginal test of $H_0 : \beta_1 = 0$ ($\chi^2_1 = 0.012$, p-value= 0.91). However, the marginal test of $H_0 : \gamma_1 = 0$ was significant ($\chi^2_1 = 4.53$, p-value= 0.033). The results indicate that having a separate pit latrine structure is not likely to increase mosquito counts to any meaningful extent.

Using these parameter estimates and based on the Wald test with expected information matrix, a new study would require sample sizes of 505, 165,000 and 419 in each group to achieve 80% power with a type-I error rate of 0.05 for the

following three tests $H_0 : [\beta_1, \gamma_1]' = [0, 0]'$, $H_0 : \beta_1: = 0$, and $H_0 : \gamma_1 = 0$, respectively. For the Wald test with Lyles, Lin, and Williamson's method with, the sample sizes calculated were 499, 163,400 and 414, respectively. The sample sizes calculated based on the likelihood ratio test were 496, 163,350 and 411. If we assumed that the mosquito count followed the usual Poisson distribution, a sample of size 323 would be required for 80% power. This example underlines the potential differences, which can be substantial, in sample size calculations based on the traditional Poisson model and the ZIP model. When the data clearly have excess zeros, it is important to use sample size calculations based on a zero-inflated model to avoid an underpowered study.

6. Discussion

Here we present power calculations for ZIP and ZINB models based on the Wald and likelihood ratio tests. If interested in a sample-size calculation instead of a power calculation, one would first specify the power and the percentage of the sample in each group. Then, by trial and error, one could calculate N_0 and N_1 using the above procedure. Example SAS programs and macros, and the malaria data set from Section 5, are available[2].

Acknowledgements

Appendix

Let $A_i = I(Y_i = 0)$ and define the probability that $A_i = 1$ as $\tau_i = \pi_i + (1 - \pi_i)e^{-\lambda_i}$. The score functions for the ZIP model are as follows:

$$\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^{N} \left\{ \tau_i^{-1} A_i \pi_i (1 - \pi_i)(1 - e^{-\lambda_i}) - \pi_i(1 - A_i) \right\} [1, x_i]'$$

$$\mathbf{S}_{\boldsymbol{\gamma}}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^{N} \left\{ -\tau_i^{-1} A_i \lambda_i e^{-\lambda_i}(1 - \pi_i) + (1 - A_i)(y_i - \lambda_i) \right\} [1, x_i]'$$

---

[2]See www.personal.psu.edu/hxl28/research/zeroinflated/data.

and the information matrix is comprised of

$$
\mathbf{I}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^{N} \pi_i (1 - \pi_i) \{ \tau_i^{-1} (1 - e^{-\lambda_i}) (pi_i^2 - e^{-\lambda_i} (1 - \pi_i)^2) + 1 - \tau_i \}
$$
$$
\times [1, x_i]'[1, x_i]
$$
$$
\mathbf{I}_{\boldsymbol{\gamma}}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^{N} \lambda_i \{ -\tau_i^{-1} (1 - \pi_i) e^{-\lambda_i} (\pi_i \lambda_i - \tau_i) + 1 - \tau_i \}
$$
$$
\times [1, x_i]'[1, x_i],
$$

and

$$
\mathbf{I}_{\boldsymbol{\beta}, \boldsymbol{\gamma}}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^{N} \tau_i^{-1} \lambda_i e^{-\lambda_i} \pi_i (1 - \pi_i) \times [1, x_i]'[1, x_i].
$$

References

Bloland, P. B., Rusbush, T. K., McCormick, J. B., Ayisi J., Borisa, D. A., Oloo, A. J., Beach, R., Hawley, W. A., Lal, A., Nahlen, B., Udhayakumar, V. and Campbell, C. C. (1999). Longitudinal cohort study of the epidemiology of malaria infections in an area of intense malaria transmission. I. Description of study site, general methodology, and study population. *American Journal of Tropical Medicine and Hygiene* **60**, 635-640.

Blom G. (1958). *Statistical Estimates and Transformed Beta-Variables*. Wiley.

Bohning, D., Dietz, E., Schlattmann, P., Mendonca, L. and Kirchner, U. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society, Series A* **162**, 195-209.

Campbell, M. J., Machin, D. and D'Arcangues, C. (1991). Coping with extra poisson variability in the analysis of factors influencing vaginal ring expulsions. *Statistics in Medicine* **10**, 241-254.

Cheung, Y. B. (2002). Zero-inflated models for regression analysis of count data: a study of growth and development. *Statistics in Medicine* **21**, 1461-1469.

Cohen, A. C. (1960). Estimating the parameters of a modified Poisson distribution. *Journal of the American Statistical Association* **55**, 139-143.

Dalrymple, M. L., Hudson, I. L. and Ford, R. P. K. (2003). Finite mixture, zero-inflated Poisson and Hurdle models with application to SIDS. *Computational Statistics and Data Analysis* **41**, 491-504.

Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics* **56**, 1030-1039.

Hall, D. B. and Berenhaut, K. S. (2002). Score tests for heterogeneity and overdispersion in zero-inflated Poisson and binomial regression models. *The Canadian Journal of Statistics* **30**, 415-430.

Krishnamoorthy, K. (2001). *Statistical Distributions with Applications and StatCalc Software.* ETEXT Textbook Publisher.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**, 1-14.

Lee, A. H., Wang, K. and Yau, K. K. W. (2001). Analysis of zero-inflated Poisson data incorporating extent of exposure. *Biometrical Journal* **43**, 963-975.

Li, C. S., Lu, J. C., Park, J., Brinkley, P. A. and Peterson, J. (1999). Multivariate zero-inflated Poisson models and their applications. *Technometrics* **41**, 29-38.

Lyles, R. H., Lin H.-M. and Williamson, J. M. A Practical approach to computing power for generalized linear models with nominal, count, or ordinal responses. To appear in *Statistics in Medicine.*

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics* **33**, 341-365.

O'Brien, R. G. (1986). Using the SAS system to perform power analyses for log-linear models. *Proceedings of the Eleventh Annual SAS Users Group International Conference*, 778-782. SAS Institute Inc.

O'Brien, R. G. (1998). A tour of UnifyPow: A SAS module/macro for sample size analysis. *Proceedings of the Twenty-Third Annual SAS Users Group International Conference* 1346-1355. SAS Institute Inc.

O'Brien, R. G. and Muller, K. E. (1993). Unified power analysis for t-tests through multivariate hypotheses. In ed., *Applied Analysis of Variance in Behavioral Science* (Edited by L. K. Edwards), 297-344. Marcel Dekker.

Ridout, M., Hinde, J. and Demetrio, G. B. (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics* **57**, 219-223.

Stroud, T. W. F. (1973). Noncentral convergence of Wald's large-sample test statistic in exponential families. *The Annals of Statistics* **1**, 161-165.

Stuart, A., Ord, J. K. and Arnold, S. (1999). *Kendall's Advanced Theory of Statistics, Sixth Edition, Volume 2A: Classical Inference and the Linear Model.* Arnold Press.

van den Broek, J. (1995). A Score test for zero inflation in a Poisson distribution. *Biometrics* **51**, 738-743.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. of the American Mathematical Society* **54**, 426-482.

Walhin, J. F. (2001). Bivariate ZIP models. *Biometrical Journal* **43**, 147-160.

Yau, K. K. W. and Lee, A.H. (2001). Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme. *Statistics in Medicine* **20**, 2907-2920.

John M. Williamson
Division of Parasitic Diseases (MS F-22)
National Center for Infectious Diseases
Centers for Disease Control and Prevention
4770 Buford Highway NE
Atlanta, Georgia 30341, USA
jow5@cdc.gov

Hung-Mo Lin
Department of Health Evaluation Sciences
Penn State College of Medicine
P. O. Box 855, A210, 600 Centerview Drive
Hershey, Pennsylvania 17033, USA
lin@psu.edu

Robert H. Lyles
Department of Biostatistics
The Rollins School of Public Health
Emory University
1518 Clifton Rd. NE
Atlanta, Georgia 30322, USA
rlyles@sph.emory.edu

Allen W. Hightower
Division of Parasitic Diseases (MS F-22)
National Center for Infectious Diseases
Centers for Disease Control and Prevention
4770 Buford Highway NE
Atlanta, Georgia 30341, USA
ahightower@ke.cdc.gov