Published in final edited form as:

Stat Methods Med Res. 2019 April ; 28(4): 1157–1169. doi:10.1177/0962280217749991.

A test of inflated zeros for Poisson regression models

Hua He¹, Hui Zhang², Peng Ye^{1,3}, and Wan Tang⁴

¹Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA

²Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA

³School of Statistics, University of International Business and Economics, Beijing, China

⁴Department of Global Biostatistics and Data Science, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA

Abstract

Excessive zeros are common in practice and may cause overdispersion and invalidate inference when fitting Poisson regression models. There is a large body of literature on zero-inflated Poisson models. However, methods for testing whether there are excessive zeros are less well developed. The Vuong test comparing a Poisson and a zero-inflated Poisson model is commonly applied in practice. However, the type I error of the test often deviates seriously from the nominal level, rendering serious doubts on the validity of the test in such applications. In this paper, we develop a new approach for testing inflated zeros under the Poisson model. Unlike the Vuong test for inflated zeros, our method does not require a zero-inflated Poisson model to perform the test. Simulation studies show that when compared with the Vuong test our approach not only better at controlling type I error rate, but also yield more power.

Keywords

Zero-inflated Poisson (ZIP); hypothesis test; Vuong test; type I error; power

1 Introduction

Count or frequency outcomes such as the number of heart attacks, suicide attempts, days of alcohol drinking, and unprotected vaginal sex occasions during a period of time arise quite often in biomedical and psychosocial research. Poisson loglinear regression is the most popular approach for modeling such count responses. However, in practice it is often the case that there are excessive zeros beyond the amount of zeros expected by the Poisson law. One common cause is the heterogeneity of the study population; subjects who are not at risk for the phenomenon of interest always yield a zero outcome, or structural zero. For example,

Reprints and permissions: sagepub.co.uk/journalsPermissions.nav

Corresponding author: Wan Tang, Department of Global Biostatistics & Data Science, Tulane University School of Public Health and Tropical Medicine, 1440 Canal Street, New Orleans, LA 70112, USA. wtang1@tulane.edu.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

when modeling behavioral outcomes such as the number of unprotected vaginal sex over a period of time in human immunodeficiency virus (HIV) prevention research, if the study population contains a subgroup of individuals who are not at risk for such health-risk behavior during the study period, every subject in this subgroup will yield a zero outcome. If a Poisson regression model is appropriate for modeling the count outcome for the at-risk subjects, the whole population follows a mixture of a Poisson distribution and a degenerate distribution at zero. A popular choice for such a mixture is the zero-inflated Poisson (ZIP) model, consisting of a Poisson regression model for the count outcome for the at-risk subjects and a regression for a binary outcome indicating the structural zero, or the nonrisk subgroup.^{1–5} The inherent methodological problems with zero-inflated outcomes have received a great deal of attention in the literature.^{6–12} However, surprisingly testing whether there are excessive zeros in the first place, a more fundamental question, has not been well investigated.

While descriptive statistics such as frequency tables or/and histograms may be informative in raising the question of whether there are excessive zeros, they should not be relied upon for decision making in place of formal statistical tests. In practice, goodness-of-fit tests may also be used to check if there are issues of fitting Poisson models, such as overdispersion. However, such tests are not really targeting the excessiveness of zeros. The Vuong test, which compares model fitting between the Poisson and a ZIP model,¹³ is widely used and has been implemented in popular statistical software packages such as SAS and Stata. A search of “zero-inflated Vuong” on Google Scholar returned about 3430 results (on 15 August 2017). However, as a general tool for comparing model fitting between two non-nested models, the Vuong test has many limitations and should not be used for testing excessive zeros in the current context.

First, the Vuong test compares the goodness of fit of the Poisson model with the ZIP model. Although not in a standard way, the Poisson is essentially nested within the ZIP model, as it corresponds to the cases where the degenerate component of the mixture (structural zeros) does not exist, that is, the probability of structural zeros is zero. Thus, it is not appropriate to apply the Vuong test for non-nested models to test inflated zeros. More specifically, the asymptotic theory which was proved under the null hypothesis that the two competing non-nested models fit the data equally well does not apply. In fact, our simulation study shows that the empirical type I error using the Vuong test may seriously deviate from the nominal type I error, confirming that it is not appropriate to apply the Vuong test for the task.

Second, a fully specific ZIP model is needed to apply the Vuong test. Building such a full-blown ZIP model can be time-consuming and may not be worth the effort, if one is only interested in testing whether there are excessive zeros under the Poisson model. Further, there may be different ways to set up the ZIP model such as using different covariates in the binary regression models for the binary indicator of structural zeros, yielding different test results.

In this paper, we develop a statistical test for excessive zeros under the Poisson model. Unlike the Vuong test, it does not require the specification of a ZIP model. The null hypothesis for the proposed test is that the Poisson model is correct and the alternative is that

there are more zeros than that are predicted by the Poisson model. In Section 2, after a brief review of the Vuong test, we develop new test statistics to test for inflated zeros under the Poisson model. Simulation studies are carried out in Section 3 to evaluate the performance of the proposed tests, with a real study example given in Section 4. The paper is concluded with a discussion in Section 5.

2 A new test of inflated zero

Consider an independent sample $(\mathbf{x}_i, y_i)$, $i = 1, \dots, n$, where y_i is a count response and $\mathbf{x}_i$ is a vector of explanatory variables. Under a Poisson log-linear regression model, we assume that the logarithm of the mean response is a linear combination of the covariates, that is

$$y_i | \mathbf{x}_i \sim \text{i.d. Poisson}(\mu_i), \quad \log(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} \quad (1)$$

In some studies, there are often excessive zeros in the count response. This may cause overdispersion and produce biased inference. Zero-inflated regression models such as ZIP model are formally used to address such violations. Under a ZIP regression model

$$y_i | \mathbf{x}_i \sim \text{i.d. ZIP}(\rho_i, \mu_i), \quad \text{logit}(\rho_i) = \mathbf{u}_i^\top \boldsymbol{\beta}_u, \quad \log(\mu_i) = \mathbf{v}_i^\top \boldsymbol{\beta}_v,$$

where $\text{ZIP}(\rho_i, \mu_i)$ is a ZIP distribution. The ZIP mixes the structural zeros with a count outcome following the Poisson distribution, with probability ρ_i from the inflated zeros and probability $(1 - \rho_i)$ from the Poisson distribution with mean μ_i . Thus, the ZIP model extends the Poisson model to account for structural zeros representing a subgroup not at risk for the phenomenon whose occurrences are captured by the Poisson model. The covariates in the two components may be different, but can have overlap.

There is an extensive literature on inference and application of ZIP and related models for zero-inflated count responses, including mixed and marginal models,^{8,9} and estimating equations.^{10,12} Zero-inflated models can correct overdispersion caused by excessive zeros and identify the nonrisk subgroup in the study population. However, a more fundamental question, which has received much less attention in the literature, is testing the existence of excessive zeros. Currently, the Vuong test¹³ is “the” test of choice and has been implemented in popular statistical software packages such as SAS and Stata. We briefly review this popular test below.

2.1 Vuong test

Let $f_1(y_i | \boldsymbol{\theta}_1)$ and $f_2(y_i | \boldsymbol{\theta}_2)$ denote the probability distribution functions of two models. If the two models fit the data equally well, then their likelihood functions would be nearly identical. Otherwise, differences between the likelihood functions provide an indication of which model fits the data better. The Vuong test is predicated upon this principle.

Let $\theta_1^*(\theta_2^*)$ be the maximum likelihood estimate (MLE) of θ_1 (θ_2) at which $\sum_{i=1}^n \log(f_1(y_i|\theta_1))(\sum_{i=1}^n \log(f_2(y_i|\theta_2)))$ achieves its maximum. The Vuong test is compare the likelihood functions at the MLE between the two models, that is,

$$\sum_{i=1}^n [\log(f_1(y_i|\theta_1^*)) - \log(f_2(y_i|\theta_2^*))] = \sum_{i=1}^n \left[\log \left(\frac{f_1(y_i|\theta_1^*)}{f_2(y_i|\theta_2^*)} \right) \right].$$

To use this ratio of likelihood functions as a test statistic, we need to find a distribution of this statistic. To this end, let

$$g_i = \log \left[\frac{f_1(y_i|\theta_1^*)}{f_2(y_i|\theta_2^*)} \right], \quad \bar{g} = \frac{1}{n} \sum_{i=1}^n g_i, \quad s_g^2 = \frac{1}{n-1} \sum_{i=1}^n (g_i - \bar{g})^2 \quad (2)$$

The Vuong test for comparing f_1 and f_2 is defined by the following statistic and associated asymptotic distribution

$$V = \frac{\sqrt{n}\bar{g}}{s_g} \sim AN(0, 1),$$

where $AN(0, 1)$ denotes the asymptotic standard normal distribution. Because of the symmetry of the two competing models, the test is nondirectional. If the absolute value $|V|$ is small, for example, the corresponding p -value is bigger than a prespecified critical value such as 0.05, then the two models fit the data equally well, with no preference given to either model. If $|V|$ yields a p -value smaller than the threshold such as 0.05, then one of the models is better; f_1 (f_2) is better if V is positive (negative). Thus, testing the null $H_0: E(g_i) = 0$ by the Vuong statistic is the same as testing a zero mean of the variable g_i .

To use the Vuong test for Poisson-distributed data, we must have a second, alternative model such as ZIP, in addition to the Poisson loglinear regression model. Often the Poisson model is treated as the null, and the alternative is the second model such as the ZIP which includes the Poisson model as a special (limiting) case. The null is rejected only if the Vuong test is significant in favor of the second model. In other words, when the Vuong test is not significant or significant but in favor of the Poisson model, we do not reject the null.

The lack of statistical tests for inflated zeros in the standard testing paradigm may be due to the fact that the two models are essentially nested, albeit in a nonstandard way. The Poisson model corresponds to the special case of ZIP model when the probability for structural zeros is zero. However, under ZIP model, a logistic model is assumed for the mixture probability, which assumes a priori the presence of structural zeros, with the proportion (probability) of such zeros inside the open interval (0,1). In other words, the Poisson model corresponds to the limiting case of ZIP model as some parameters go to negative infinity, rather than a member of the ZIP model in conventional sense. For example, if there is no covariate in the logistic component of ZIP model, then the Poisson model corresponds to the case when the

intercept is negative infinite. The issue is more complicated if there are covariates in the model. There are more than one parameter in the logistic component and the limiting situation can become quite complicated. Thus, fitting a second model such as the ZIP when using the Vuong test not only involves additional more complex modeling (than Poisson model), but may suffer convergence issue as well, especially when the probabilities of being structural zeros is small.

2.2 A naive test

Intuitively, one may compare the amount of zeros observed with that expected under the Poisson regression model. Let r_i be the indicator of $y_i \neq 0$, that is, $r_i = 1 (= 0)$ if $y_i = 0 (>0)$. Under equation (1), the probability that $y_i \neq 0$ with covariate $\mathbf{x}_i$ is $\exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta}))$. Thus, the expected proportion of zero outcomes is

$$E(r_i) = \frac{1}{n} \sum_{i=1}^n \exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta})),$$

and the difference between observed and expected proportions of zero outcomes is

$$s = \frac{1}{n} \sum_{i=1}^n (r_i - E(r_i)) = \frac{1}{n} \sum_{i=1}^n r_i - \frac{1}{n} \sum_{i=1}^n \exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta})) \quad (3)$$

Under equation (1), $E(s)=0$ and $Var(s) = \frac{1}{n^2} \sum_{i=1}^n \exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta}))(1 - \exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta})))$, so the normalized version of s , s' , is asymptotically normal

$$s' = \frac{\frac{1}{n} \sum_{i=1}^n (r_i - \exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta})))}{\left[\frac{1}{n^2} \sum_{i=1}^n \exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta}))(1 - \exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta}))) \right]^{1/2}} \sim AN(0, 1) \quad (4)$$

However, neither s nor s' is in general a statistic because usually the parameter $\boldsymbol{\beta}$ in equation (1) is unknown. A naive approach would be to first estimate $\boldsymbol{\beta}$ and then use equation (3) or (4) after replacing $\boldsymbol{\beta}$ with a consistent estimate.

The parameter $\boldsymbol{\beta}$ can be estimated using maximum likelihood, or equivalently, by solving the score equations

$$\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^T (y_i - \exp(\mathbf{x}_i^T \boldsymbol{\beta})) = 0 \quad (5)$$

Let $\hat{\boldsymbol{\beta}}$ be the MLE. By substituting $\hat{\boldsymbol{\beta}}$ in place of $\boldsymbol{\beta}$ in equation (3) or (4), we obtain an estimate of s or s'

$$\begin{aligned}\hat{s} &= \frac{1}{n} \sum_{i=1}^n r_i - \frac{1}{n} \sum_{i=1}^n \left(\exp(-\exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})) \right), \\ \hat{s}' &= \frac{\frac{1}{n} \sum_{i=1}^n (r_i - \exp(-\exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})))}{\frac{1}{n} \left[\sum_{i=1}^n \exp(-\exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})) (1 - \exp(-\exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}))) \right]^{1/2}}\end{aligned}\quad (6)$$

Inference based on this naive statistic against the asymptotic standard normal will be biased because it treats the estimated $\hat{\boldsymbol{\beta}}$ as true $\boldsymbol{\beta}$, thereby ignoring completely the sampling variation associated with $\hat{\boldsymbol{\beta}}$. Next, we describe how to correct this bias for valid inference.

2.3 A new approach

For a valid statistical test based on equation (3), we need to correctly handle the variation associated with the estimated $\boldsymbol{\beta}$. To this end, first note that equation (3) can be expressed as an estimate of $S = E(s) = E(r_i - \exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta})))$, the mean excessive probability of being a zero over that expected under a Poisson model, by solving the following estimating equation

$$\frac{1}{n} \sum_{i=1}^n (r_i - \exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta})) - S) = 0 \quad (7)$$

By stacking equation (7) with the estimating equations for $\boldsymbol{\beta}$ (equation (5)), we obtain a system of estimating equations which simultaneously solve $\boldsymbol{\beta}$ and S . It is easy to check that the solution for S is exactly the same as $\hat{s}$ in equation (6); however, the point is that the correct asymptotic distribution of $\hat{S}$ is then provided by the theory of estimating equations.

Specifically, let

$$\begin{aligned}\psi_1 &= r_i - \exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta})) - S, \\ \psi_2 &= \mathbf{x}_i^T (y_i - \exp(\mathbf{x}_i^T \boldsymbol{\beta})), \\ A &= E \begin{pmatrix} \frac{\partial \psi_1}{\partial s} & \frac{\partial \psi_1}{\partial \boldsymbol{\beta}} \\ \frac{\partial \psi_2}{\partial s} & \frac{\partial \psi_2}{\partial \boldsymbol{\beta}} \end{pmatrix} = E \begin{pmatrix} -1 & \mathbf{x}_i^T \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta})) \\ 0 & -\exp(\mathbf{x}_i^T \boldsymbol{\beta}) \mathbf{x}_i \mathbf{x}_i^T \end{pmatrix}, \\ B &= E \begin{pmatrix} \psi_1^2 & \psi_1 \psi_2^T \\ \psi_1 \psi_2 & \psi_2 \psi_2^T \end{pmatrix}.\end{aligned}$$

The following theorem provides the asymptotic distribution for $\hat{S}$.

Theorem 1. *Under the null hypothesis of the Poisson model in equation (1), we have that $\hat{S}$ is asymptotically normal. More precisely*

$$\sqrt{n}(\hat{s} - 0) \rightarrow N(0, \sigma^2),$$

where σ^2 is the (1,1) term of $A^{-1}BA^{-T}$ and it can be estimated from sample moments of A and B by plugging in the estimations $\hat{\beta}$ and $\hat{S}$ for β and S .

Proof. Under the null hypothesis of the Poisson model in equation (1), the estimating equations for (S, β) ,

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n (r_i - \exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta}))) - S = 0 \\ \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^T (y_i - \exp(\mathbf{x}_i^T \boldsymbol{\beta})) = 0, \end{cases}$$

are unbiased, it follows from the theory of estimating equations that the estimate obtained as the solution to the estimating equations is consistent and asymptotically normal

$$\sqrt{n}((\hat{s}, \hat{\beta}) - (S, \beta)) \rightarrow N(\mathbf{0}, \Sigma),$$

where $\Sigma = A^{-1}BA^{-T}$. In particular

$$\sqrt{n}(\hat{s} - S) \rightarrow N(0, \sigma^2) \quad (8)$$

where σ^2 denotes the (1,1) term of Σ . Note that $S = 0$ under the null hypothesis, the proof is completed.

Based on this theorem, we can test excessive zeros simply based on the null hypothesis of the Poisson model, without assuming any alternative model such as the ZIP as in applying the Vuong test. If we test against the alternative in which the amount of zeros is different from that expected by the Poisson model in either direction, a two-sided test based on $\hat{s}$ can be used. For example, for a two-sided test with type I error 0.05, we reject the null hypothesis if

$$\left| \frac{\hat{s}}{\hat{\sigma}} \right| > 1.96,$$

where $\hat{\sigma}^2$ is the asymptotic variance in equation (8) obtained by plugging in the estimates of β and S .

In practice, we often test against the alternative with excessive zeros and may use a one-sided test to have more power. In this case, the alternative is in the direction of excessive zeros beyond that of the Poisson model and we reject the null if

$$\frac{\hat{s}}{\hat{\sigma}} > 1.64.$$

3 Simulation studies

We conduct simulation studies to assess the performance of the proposed test. We simulate data in different scenarios (Poisson distribution and regression models, ZIP models with linear and nonlinear components, and negative binomial (NB) models) with different sample sizes (50, 100, 200, 500, and 1000) to compare the new, naive, and the Vuong tests. We first simulate data from Poisson models to assess the performances in terms of type I errors. Next, we simulate data from ZIP models to assess the performance in terms of power. We simulate data using constant, linear, and nonlinear models for the structural zero components. Finally, we simulate data following NB models. The NB has the same first moment as the Poisson distribution, but a larger variance than the Poisson distribution.

All the tests are one-sided unless stated otherwise. For our new and naive tests, the alternative is that there are more zeros than those predicted by the Poisson model. For Vuong test, the alternative is the ZIP model with the same Poisson model for the count component and a logistic model for the structural zero component. All simulations are performed with 10,000 Monte Carlo replicates and the statistical significance level at $\alpha = 0.05$.

The simulation is carried out using R.¹⁴ R programs are developed for the new test, and they are available upon request. ZIP models are fitted using the “zeroinfl” function and the Vuong statistics are computed using the “vuong” function in the R package “pscl”.¹⁵ Note that the original “vuong” function outputs the test results, but we have made simple changes to output the Vuong statistic.

3.1 Poisson response

We consider two situations, with the first (second) involving no (one) covariate. We express both scenarios in one model as follows

$$y_i | x_i \sim \text{Poisson}(\mu), \mu = \exp(\beta_0 + \beta_1 x_i), \quad x_i \sim N(0, 1).$$

The two situations correspond to $\beta_1 = 0$ (first) and 1 (second).

For the first scenario, we simulate data from Poisson distributions with mean $\mu = 0.5$ and 1, so the corresponding $\beta_0 = \log(0.5)$ and 0. The null hypothesis is

$$H_0: y \sim \text{Poisson}(\mu), \quad \mu = \exp(\xi_0) \quad (9)$$

and the ZIP model for the Vuong test is

$$y \sim \text{ZIP}(\mu, \rho), \mu = \exp(\eta_0), \quad \text{logit}(\rho) = \gamma_0,$$

where μ and ρ are constants.

For the second case with covariate x , we set $\beta_0 = 0.5$ and 1. The null hypothesis is

$$H_0: y \sim \text{Poisson}(\mu), \quad \mu = \exp(\xi_0 + \xi_1 x).$$

For the Vuong test, we consider two alternative ZIP models, the first with a constant mixing probability

$$H_{a1}: y \sim \text{ZIP}(\mu, \rho), \quad \mu = \exp(\eta_0 + \eta_1 x), \quad \text{logit}(\rho) = \gamma_0 \quad (10)$$

and the second including the covariate in the component for inflated zeros

$$H_{a2}: y \sim \text{ZIP}(\mu, \rho), \quad \mu = \exp(\eta_0 + \eta_1 x), \quad \text{logit}(\rho) = \gamma_0 + \gamma_1 x \quad (11)$$

Note that the null (alternative) hypothesis is the form of a model within the current context, rather than specific values of parameters as in testing hypotheses concerning the values of parameters.

Summarized in Table 1 are the empirical type I errors for the two scenarios. In all cases, the new test showed consistently good performances; the deviations of empirical type I errors from the nominal 0.05 lied within the simulation margin.¹⁶ On the other hand, the naive and the Vuong tests are quite problematic. As expected, the naive test rejected the hypothesis at a much lower rate than the nominal level. The Vuong test did not perform well in all the cases either, with bias in both directions. In the first scenario, the Vuong test had a much lower reject rate than the nominal 5% level, whereas in the second scenario, it had a much lower (higher) reject rate than the nominal 5% level if the covariate is excluded (included) in the structural zero component of the ZIP model.

Since the Poisson model is essentially nested within the ZIP model, the Vuong test comparing the goodness of fit of the two models in general favors the ZIP model over the Poisson regression model. The degree to which it favors the ZIP over the Poisson model varies with different situations. When there are no covariates in the model for the structural zero component, the advantage of the ZIP over the Poisson model seems minimal and their difference in likelihood is small, resulting in rejection rates that are close to zero (all less than 0.15%). However, when covariates are included in the model for the structural zero component, more parameters are added to the ZIP model and the additional parameters bring more flexibility for the ZIP model to fit the data and may yield a substantially higher increase in the likelihood for the ZIP model. In fact, by simply switching the ZIP model from not including the covariate to including the covariate, the rejection rates change from close to zero to much higher than the nominal 5% type I error, neither of which is what we desired. The substantial deviations of the empirical type I errors from the nominal level in both directions prove that the Vuong test totally failed in controlling type I error, invalidating its application in testing inflated zeros.

Shown in Figure 1 are QQ plots of the p -values for comparing model-based and empirical type I errors for the simulation study; the plots in the four rows represent the four cases in

Table 1. Within each row, the plots correspond to the Vuong, naive, and new test, respectively. For the Vuong test in the third and fourth rows in the second scenario, the curves above (below) the diagonal lines are for the test using ZIP models without (with) the covariate in the structural zero component. The plots also show that the new test performed quite well, because the points showed near perfect diagonal lines. On the other hand, the QQ plots for the naive and the Vuong test show quite deviations from the diagonal line. For all cases, the bottom-left end of the plot for the naive test lied above the diagonal line, indicating that the model-based p -values were far bigger their empirical ones, hence resulting in much lower rejection rates. For the Vuong test, the plots showed similar patterns for the two cases in the first scenario as well as in the second scenario when the covariate is not included in the structural zero component of the ZIP model. But, when the covariate is included in the zero component of the ZIP model in the second scenario, the points all lied below the diagonal line, indicating that the estimated p -values were smaller than the empirical ones, leading to much higher rejection rates.

3.2 ZIP response

Next, we simulate responses from the ZIP model and focus on the performance of the tests in terms of power. Again, we assume a single explanatory variable. The Poisson component of the ZIP model is given by

$$y \sim \text{Poisson}(\mu), \quad \mu = \exp(a + x), \quad x_i \sim N(0, 1).$$

For the logistic part (mixture probability), we simulate data from three scenarios; (1) a logistic regression with only the intercept (constant mixing probability); (2) a logistic model with a linear predictor; and (3) a logistic model with a nonlinear predictor.

The null considered is the Poisson model

$$y \sim \text{Poisson}(\mu), \quad \mu = \exp(\beta_0 + \beta_1 x).$$

For the Vuong test, we need to specify a ZIP model. We use the following ZIP model

$$y \sim \text{ZIP}(\mu, \rho), \quad \mu = \exp(\beta_0 + \beta_1 x), \quad \text{logit}(\rho) = \gamma_0 + \gamma_1 x,$$

for all the cases, even for the nonlinear cases, as this is normally assumed in practice.

3.2.1 Constant mixing probability—In this case, all subjects have the same likelihood of being a structural zero, ρ . So, it reduces to the Poisson case if $\rho = 0$. Here, we simulate data using $\rho = 0.05, 0.1, 0.15$ and 0.2 according to the following ZIP model

$$y \sim \text{ZIP}(\mu, \rho), \quad \mu = \exp(a + bx), \quad \rho = 0.05, 0.1, 0.15, 0.2.$$

Summarized in Table 2 are the empirical power (proportion of correctly rejecting the null) for each of the tests. As ρ increases, the amount of structural zeros and thus power should

also increase. This pattern is quite clear from the results for all the three methods. The Vuong test does not show any advantage over the proposed test. In fact, in most cases the new test is more powerful.

3.2.2 Logistic model with a linear predictor mixing probability—The mixing probability in this case follows a generalized linear model with the logit link and a single covariate, $\text{logit}(\rho) = \alpha + x$. The response is generated according to the following model

$$y \sim \text{ZIP}(\mu, \rho), \quad \mu = \exp(x), \quad \text{logit}(\rho) = \alpha + x, \quad \alpha = -2.94, -2.20, -1.73, -1.39.$$

The different values of α lead to approximately 10%, 15%, 20%, and 25% of structural zeros.

Shown in Table 3 are the empirical power (proportion of correctly rejecting the null) for the three tests corresponding to the range of α considered. Again, power increases across the board as α or the mixing probability increases. As before, the Vuong test does not perform better than the new test.

3.2.3 Logistic model with a nonlinear predictor as mixing probability—The mixing probability here still follows a logistic model, but linking to the covariate in a nonlinear fashion, $\text{logit}(\rho) = \alpha |\sin(4\pi x)|$, where $\alpha = 0.1, 0.2, 0.3$, and 0.4 . The data are generated according to the following model

$$y \sim \text{ZIP}(\mu, \rho), \quad \mu = \exp(x), \quad \text{logit}(\rho) = \alpha |\sin(4\pi x)|.$$

As in the earlier case, the mixing probability increases if α increases.

For the Vuong test, ZIP models with generalized linear models for both components (equation (10)) are used as the alternative.

Summarized in Table 4 are the empirical power for each of the tests. Once again, the power increases for all three methods as α increases. Compared with the proposed test, the Vuong test has lower power values in most of the cases.

3.3 NB response

Data in this section are simulated from an NB regression model: $y \sim \text{NB}(\mu, \tau)$, where the distribution function is given by

$$f_{\text{NB}}(y|\mu, \tau) = \frac{\Gamma(y + \tau)}{y! \Gamma(\tau)} \left(\frac{1}{1 + \mu/\tau} \right)^y \left(\frac{\mu/\tau}{1 + \mu/\tau} \right)^\tau, \quad \tau > 0, \quad y = 0, 1, 2, \dots \quad (12)$$

In the simulation, we set $\tau = 1$ and $\log(\mu) = \alpha x$, with $\alpha = 0, \frac{1}{3}, \frac{2}{3}$ and 1 . The null hypothesis is still the Poisson model (equation (9)) and the following ZIP model is used for the Vuong test

$$y \sim \text{ZIP}(\mu, \rho), \quad \mu = \exp(a + bx), \quad \text{logit}(\rho) = a + bx.$$

The rejection rates for all of the tests are shown in Table 5. For sample sizes of 50 and 100, the new test had considerably higher rejection rates than the Vuong test. For sample sizes of 200, 500, and 1000, their rejection rates were comparable. As expected, the rejection rate increased with the sample size in all cases for all the tests.

Note that the rejection of the null by the Vuong test indicates that the ZIP model fits the data better than the Poisson model, while the rejection by the new test shows that there are more zeros than those predicted by the Poisson model. Within the current simulation context, the latter is true, because there are more zeros in an NB than in a Poisson model, if the two models have the same mean. However, the rejection of the null hypothesis here does not imply inflated zeros or existence of structural zeros. The new test only compares the amount of zeros with that expected by the Poisson model. Thus, whether the inflated zeros detected by the proposed test can be interpreted as the indication of the existence of the structural zeros depends on whether the at-risk group follows the Poisson model, which may be checked by comparing the positive count outcome with the corresponding zero-truncated Poisson model.

4 Case study

In a randomized clinical trial, teaching awareness and self-monitoring skills to indwelling urinary catheter users conducted in New York state, 202 subjects were recruited and randomized to the intervention and control groups.¹⁷ The primary outcomes of interest are whether the subjects experienced urinary tract infections (UTIs), catheter blockages, and catheter displacements during the last two months, as well as the corresponding counts of these experiences. This is a randomized longitudinal study where each subject was measured every 2 months, however, for illustrative purpose, we only consider the count responses at enrollment.

We would like to apply Poisson regression models to model the count responses. At the enrollment, the distributions of the count outcomes are presented in the following plot.

Based on the plot, it looks like that there may be the zero-inflation issue for these count outcomes. To formally test if it is indeed the case, we use the following Poisson regression model

$$y_i | x_i \sim \text{Poisson}(\mu_i), \quad \mu_i = \exp(\beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Gender}_i + \beta_3 \text{Time}_i + \beta_4 \text{Type}_i) \quad (13)$$

where y_i is the count response (count of UTIs, count of catheter blockages, and count of catheter displacements) and the variables age, gender, time, and type are the age and gender of the subjects, the duration the subject had used catheter consistently (in months), and the types of catheter used. Two types of catheters were used by all the subjects, and there was one subject used both, so this subject is excluded from the analysis.

From Figure 2, there are three measures in catheter blockages that were obvious too large (outliers). For illustrative purpose, we simply delete these measures in the analysis, although more complicated methods such as winsorizing may be applied. Based on the Poisson

regression models (equation (13)) for the three outcomes, the p -values for the tests of zero-inflation based on our proposed methods are 0.68, 0.0000015, and 0.0016, respectively.

To apply the Vuong test, we need to set up an alternative ZIP regression model. We assume the Poisson regression models (equation (13)) for the count component. For the binary component for inflated zeros, there are $2^4 = 16$ different choices, even only main effects are considered. To save space, we consider the following eight models:

1. $y_i \sim \text{ZIP}(\mu_i, \rho_i)$, $\text{logit}(\rho_i) = \gamma_0$, there are no predictors in the binary component.
2. $y_i \sim \text{ZIP}(\mu_i, \rho_i)$, $\text{logit}(\rho_i) = \gamma_0 + \gamma_1 \text{Age}_i$
3. $y_i \sim \text{ZIP}(\mu_i, \rho_i)$, $\text{logit}(\rho_i) = \gamma_0 + \gamma_1 \text{Gender}_i$
4. $y_i \sim \text{ZIP}(\mu_i, \rho_i)$, $\text{logit}(\rho_i) = \gamma_0 + \gamma_1 \text{Type}_i$
5. $y_i \sim \text{ZIP}(\mu_i, \rho_i)$, $\text{logit}(\rho_i) = \gamma_0 + \gamma_1 \text{Time}_i$
6. $y_i \sim \text{ZIP}(\mu_i, \rho_i)$, $\text{logit}(\rho_i) = \gamma_0 + \gamma_1 \text{Age}_i + \gamma_2 \text{Gender}_i$
7. $y_i \sim \text{ZIP}(\mu_i, \rho_i)$, $\text{logit}(\rho_i) = \gamma_0 + \gamma_1 \text{Age}_i + \gamma_2 \text{Gender}_i + \gamma_3 \text{Type}_i$
8. $y_i \sim \text{ZIP}(\mu_i, \rho_i)$, $\text{logit}(\rho_i) = \gamma_0 + \gamma_1 \text{Age}_i + \gamma_2 \text{Gender}_i + \gamma_3 \text{Type}_i + \gamma_4 \text{Time}_i$

For UTI, the corresponding p -values under the Vuong tests are 0.37, 0.39, 0.44, 0.49, 0.42, 0.04, 0.10, and 0.084, respectively, for the eight ZIP models.

For catheter blockages, the corresponding p -values under the Vuong tests are 0.0015, 0.0015, 0.00098, 0.00098, 0.0015, 0.0010, 0.0010, and 0.00054, respectively, for the eight ZIP models.

For catheter displacement, the corresponding p -values under the Vuong tests are 0.043, 0.043, 0.044, 0.033, 0.030, 0.044, 0.012, and 0.012, respectively, for the eight ZIP models.

Based on these tests, the p -values for the Vuong tests vary with the model specification for the zero component. For UTI, they vary a lot, from 0.04 to 0.49. Different conclusions may be achieved with 5% type I error. Similarly, the p -values change with the different models applied for the zero-component for catheter blockages and displacement. Different conclusions may be resulted with 0.1% type I error for catheter displacement.

5 Discussion

Inflated zeros is a ubiquitous issue in biomedical and psychosocial research and practice. Although an extensive literature exists on zero-inflated models and their applications, testing of inflated zeros is underdeveloped and the Vuong test is “the” test for such purposes. However, the Vuong test is not appropriate for testing inflated zeros, because of the serious problem with type I error. The Poisson model is essentially nested in the ZIP model and hence the asymptotic theory of the Vuong statistic for non-nested models is no longer valid. As our simulation study shows, the empirical type I error of the Vuong test may seriously deviate from the nominal one in either directions, even when the sample size is large.

In this paper, we developed a new statistic to address the gap in the current literature. Unlike the Vuong test, the proposed approach specifically tests inflated zeros, with straightforward interpretations. Asymptotic properties for our new test are developed for large samples, and simulation studies show that it also performs quite well, in terms of both controlling type I error and power, for small to moderate sample sizes. In all the cases considered, our test shows much better performance in controlling type I error and similar or better performance in power than the Vuong test. Further, unlike the Vuong test, there is no need to specify a ZIP model to perform the test. Thus, our method avoided the uncertainty associated with the specification of different ZIP models.

The new test is based on a Poisson model, so different conclusions may still be obtained if different Poisson models are utilized. This seems unavoidable since the phenomenon of zero-inflation refers to the fact that there are more zeros than that would be expected under some models such as Poisson. Note that our test does not test the validity of the Poisson model. Thus, ideally, the test may be applied when a Poisson model is validated for the at-risk group. For example, one may apply a goodness-of-fit test for the corresponding zero-truncated Poisson model to the data with all zero responses deleted.

We only considered the Poisson model in this paper. The approach may be extended to other distributions. Although analytic forms may vary, the same considerations should apply when considering other distributions.

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by NIH grants R33 DA027521, R01GM108337, and P20GM109036.

References

1. Calsyn DA, Hatch-Maillette M, Tross S, et al. Motivational and skills training HIV/sexually transmitted infection sexual risk reduction groups for men. *J Subst Abuse Treat* 2009; 37: 138–150. [PubMed: 19150206]
2. Crits-Christoph P, Gallop R, Sadicario JS, et al. Predictors and moderators of outcomes of HIV/STD sex risk reduction interventions in substance abuse treatment programs: a pooled analysis of two randomized controlled trials. *Subst Abuse Treat Prev Policy* 2014; 9: 3. [PubMed: 24433412]
3. Gurmu S and Trivedi PK. Excess zeros in count models for recreational trips. *J Bus Econ Stat* 1996; 14: 469–477.
4. Hur K, Hedeker D, Henderson W, et al. Modeling clustered count data with excess zeros in health care outcomes research. *Health Serv Outcomes Res Methodol* 2002; 3: 5–20.
5. Yau KK and Lee AH. Zero-inflated poisson regression with random effects to evaluate an occupational injury prevention programme. *Stat Med* 2001; 20: 2907–2920. [PubMed: 11568948]
6. Lambert D Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics* 1992; 34: 1–14.
7. Cheung YB. Zero-inflated models for regression analysis of count data: a study of growth and development. *Stat Med* 2002; 21: 1461–1469. [PubMed: 12185896]
8. Hall DB. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics* 2000; 56: 1030–1039. [PubMed: 11129458]
9. Hall DB and Zhang Z. Marginal models for zero inflated clustered data. *Stat Model* 2004; 4: 161–180.

10. Yu Q, Chen R, Tang W, et al. Distribution-free models for longitudinal count responses with overdispersion and structural zeros. *Stat Med* 2013; 32: 2390–2405. [PubMed: 23239019]
11. Tang W, He H and Tu X. *Applied categorical and count data analysis*. Boca Raton, FL: Chapman & Hall/CRC, 2012.
12. Tang W, Lu N, Chen T, et al. On performance of parametric and distribution-free models for zero-inflated and over-dispersed count responses. *Stat Med* 2015; 34: 3235–3245. [PubMed: 26078035]
13. Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 1989; 57: 307–333.
14. R Development Core Team. *R: a language and environment for statistical computing* Vienna, Austria: R Foundation for Statistical Computing, 2011 ISBN 3–900051–07–0.
15. Jackman S *pscl: Classes and Methods for R Developed in the Political Science*. Stanford, California: Computational Laboratory, Stanford University. Department of Political Science, Stanford University.
16. le Cessie S and van Houwelingen HC. Testing the fit of a regression model via score tests in random effects models. *Biometrics* 1995; 51: 600–614. [PubMed: 7662848]
17. Wilde MH, McMahon JM, McDonald MV, et al. Self-management intervention for long-term indwelling urinary catheter users: randomized clinical trial. *Nurs Res* 2015; 64: 24. [PubMed: 25502058]

Figure 1.

QQ plots of p -values for comparing model-based and empirical type I errors in the four Poisson response cases with sample sizes 50, 100, 200, 500, and 1000. For the Vuong test in the last two rows, the curves above (below) the diagonal lines are for the test using zero-inflated Poisson models without (with) the covariate in the structural zero component.

Figure 2. Distributions of the counts of urinary tract infection, catheter blockages, and catheter displacements.

Table 1.

Empirical type I errors (%) for comparing three methods under different scenarios for Poisson response for the simulation study.

Sample Size	First scenario ($\beta_1 = 0$)						Second scenario ($\beta_1 = 1$)							
	$\mu = 0.5$			$\mu = 1$			$\beta_0 = 0.5$				$\beta_0 = 1$			
	Vuong	Naive	New	Vuong	Naive	New	Vuong1	Vuong2	Naive	New	Vuong1	Vuong2	Naive	New
50	0.09	0.05	3.82	0.13	0.60	4.81	0.03	12.84	0.40	4.13	0	26.27	0.50	4.59
100	0.02	0.06	3.95	0.05	0.53	4.64	0.03	13.75	0.40	3.68	0	26.82	0.49	4.94
200	0.05	0.03	4.42	0.05	0.45	4.55	0.01	13.05	0.29	4.06	0	24.64	0.49	4.48
500	0.08	0.06	4.47	0.04	0.49	4.77	0.03	11.14	0.49	4.78	0	21.49	0.73	4.86
1000	0.03	0.03	4.43	0.04	0.54	4.51	0.03	10.02	0.52	4.74	0	17.60	0.65	5.34

Note: In the second scenario, “Vuong1” and “Vuong2” are for the Vuong tests with the alternative zero-inflated Poisson models (3.10) and (3.11), respectively.

Table 2.

Power (%) for comparing three methods under different scenarios for zero-inflated Poisson response with constant mixing probability for the simulation study.

Sample Size	$\rho = 0.05$			$\rho = 0.1$			$\rho = 0.15$			$\rho = 0.2$		
	Vuong	Naive	New	Vuong	Naive	New	Vuong	Naive	New	Vuong	Naive	New
50	11.50	12.70	19.10	11.20	26.80	40.00	22.00	43.60	64.40	39.00	65.00	85.70
100	9.20	20.90	27.60	23.90	48.70	66.50	48.70	71.90	88.10	77.70	89.00	97.60
200	11.00	32.10	42.50	58.10	70.30	86.60	89.80	91.90	98.30	98.90	99.10	99.90
500	54.70	57.70	72.80	98.50	96.50	99.40	100.00	99.90	100.00	100.00	100.00	100.00
1000	91.00	83.60	92.50	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 3.

Power (%) for comparing three methods under different scenarios for zero-inflated Poisson response with mixing probability modeled by logistic regression with a linear predictor for the simulation study.

Sample Size	$\alpha = -2.94$			$\alpha = -2.20$			$\alpha = -1.73$			$\alpha = -1.39$		
	Vuong	Naive	New	Vuong	Naive	New	Vuong	Naive	New	Vuong	Naive	New
50	23.84	12.05	26.65	45.71	27.37	49.85	63.16	39.09	66.03	76.19	48.57	77.37
100	40.04	25.50	43.14	79.01	56.63	76.24	93.95	76.67	91.25	98.12	86.36	96.51
200	76.38	49.92	66.83	98.89	88.36	95.12	99.94	97.87	99.48	99.99	99.45	99.90
500	99.75	88.62	94.76	100.00	99.85	99.93	100.00	100.00	99.99	100.00	100.00	100.00
1000	100.00	99.59	99.85	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 4.

Power (%) for comparing three methods under different scenarios for zero-inflated Poisson response with mixing probability modeled by logistic regression with a nonlinear predictor for the simulation study.

Sample Size	$\alpha = -2.94$			$\alpha = -2.20$			$\alpha = -1.73$			$\alpha = -1.39$		
	Vuong	Naive	New	Vuong	Naive	New	Vuong	Naive	New	Vuong	Naive	New
50	16.25	2.93	12.92	13.55	8.89	25.47	19.54	18.19	42.35	30.93	29.11	58.51
100	11.35	5.26	18.32	17.14	18.82	41.67	38.43	39.36	65.93	62.82	60.71	84.10
200	10.50	10.12	28.71	40.46	38.72	65.59	77.78	71.20	89.17	95.30	91.66	98.19
500	31.41	26.10	52.89	92.31	81.58	94.10	99.82	98.71	99.85	100.00	99.99	100.00
1000	76.86	53.02	78.24	99.87	98.65	99.80	100.00	100.00	100.00	100.00	100.00	100.00

Table 5.

Power (%) for comparing three methods under different scenarios for negative binomial response for the simulation study.

Sample Size	$\alpha = 0$			$\alpha = 1/3$			$\alpha = 2/3$			$\alpha = 1$		
	Vuong	Naive	New	Vuong	Naive	New	Vuong	Naive	New	Vuong	Naive	New
50	27.24	55.01	75.33	27.33	56.62	76.15	27.19	58.80	75.24	26.37	61.00	71.29
100	61.21	87.48	96.21	63.16	89.42	96.91	66.47	90.92	96.06	64.95	87.86	89.92
200	95.97	99.52	99.93	96.87	99.72	99.96	97.82	99.66	99.82	97.21	98.51	97.54
500	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.95	99.64
1000	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.89