Group Comparisons Involving Zero-Inflated Count Data in Clinical Trials

Kanae Togo\textsuperscript{*1} and Manabu Iwasaki\textsuperscript{*2}
\textsuperscript{*1}Clinical Statistics, Pfizer Japan Inc.
\textsuperscript{*2}Department of Computer and Information Science, Seikei University
\textit{e-mail:} Kanae.togo@pfizer.com

In clinical trials, outcomes of count data sometimes have excess zeros. When a test drug is compared to a control, zero-inflated data may be ignored or interest is taken only in the proportion of zero counts. By applying the two-part model, Lachenbruch (2001a) suggested a test statistic called the two-part statistic that combines the test statistics of the zero part and the non-zero part. The test for the zero part is the chi-square test. The test for the non-zero part may be a Wilcoxon test, a $t$-test, etc. This article proposes methods for calculating the sample size and power for the two-part statistic with zero-inflated Poisson data. We developed the methods of sample size and power for the two-part statistic using the Wilcoxon test adjusted for ties. The relationship between the non-zero part and zero-truncated Poisson distribution is also described. Furthermore, we examine the power of the two-part statistic, conventional methods, and the zero-inflated Poisson model.

\textit{Key words:} two-part statistic, zero-inflated Poisson model, Wilcoxon test adjusted for ties, sample size, power.

1. Introduction

In clinical trials, outcomes of count data sometimes have excess zeros. These outcomes may be found in the number of symptoms (e.g., urinary incontinence episodes, gastrointestinal ulcers, hot flushes arising from menopausal disorder), the number of events (e.g., hospitalizations, heart attacks), and questionnaire scores. When zero values are observed after treatment in clinical trials to evaluate treatment effect, it represents two types of outcome: that in which patients recover, and that in which patients do not recover but have a small value of the outcome that is zero by chance. The zero-inflated Poisson (ZIP) distribution or zero-inflated negative binomial distribution can be applied to count data with excess zeros. This article focuses on the ZIP distribution. The ZIP distribution has two parameters; $\lambda$ is the Poisson parameter and $\omega$ expresses the extent of zero-inflation compared with zero counts that occur from the Poisson distribution.

Received March 2013. Revised September 2013. Accepted October 2013.
When the treatment difference between a test drug and a control is tested, either zero-inflation or the difference in the non-zero part is sometimes ignored. For example, a nonparametric test may be used after applying rank transformation to the data. This poses the problem that there can be many ties due to zero-inflation. In this case, the normal approximation is not accurate, whereas many nonparametric tests use the normal approximation for rank data. In addition, the power to detect the treatment difference could be decreased if there are many ties in the two treatment groups. Another example is that the treatment groups are compared only in the proportion of subjects with zero as a dichotomous response. This possibly wastes the treatment difference in the non-zero part by ignoring that.

In contrast, there are two methods to compare treatment groups considering zero-inflated counts. One method is the zero-inflated count model (Lambert, 1992), and the other is the two-part model (Heilbron, 1992). The zero-inflated count model is a mixture model of a logit model for \( \omega \) and a Poisson regression model for \( \lambda \). The treatment effect is assessed by including a covariate for the treatment group in the logit and Poisson regression models. In clinical trials, however, it may be undesirable to assess the treatment effect by performing two hypothesis tests for each model because of the difficulty in interpreting the two possibly inconsistent test results and controlling the type I error inflation.

The two parts of the two-part model are defined as the zero part, consisting of the response dichotomized as zero vs. non-zero, and the non-zero part, consisting of the non-zero counts. The response variable follows the binominal distribution in the zero part and the zero-truncated count distribution in the non-zero part. By applying the two-part model, Lachenbruch (2001a) proposed a test statistic called the two-part statistic that combines the test statistics of the zero and non-zero parts. The test for the zero part is generally the chi-square test. Possible tests for the non-zero part are the Wilcoxon test, \( t \)-test, etc. This two-part test statistic provides a comprehensive result for both the zero and non-zero parts. Delucchi et al. (2004) demonstrated the two-part statistic with zero-inflated counts in the Addiction Severity Index in heroin addicts. As an alternative to the two-part statistic, Hallstorm (2010) proposed a modified Wilcoxon test for zero-inflated data that discards an equal number of zeros in each group.

This article provides methods for finding the sample size and power for the two-part statistic. Lachenbruch (2001b) developed a method for calculating the sample size when the test for the non-zero part was a \( t \)-test. Given the considerable ties in count data, however, Lanchenbruch’s method can underestimate the sample size when the Wilcoxon test is used for the non-zero part. We propose methods that, by adjusting for ties, calculate the sample size and power for the two-part statistic when the Wilcoxon test is used for the non-zero part. Furthermore, we examine, under three cases, the power of the two-part statistic compared with the conventional methods and the ZIP model. There is treatment difference in the zero-inflation \( \omega \) under the first of these cases, treatment difference in the Poisson parameter \( \lambda \) under the second, and treatment
difference in both parameters under the third. Section 2 presents the methods for calculating the sample size and the power using the two-part statistic. The relationship between the non-zero part and the zero-truncated Poisson distribution is also described. In Section 3, the power of the two-part statistic introduced in Section 2 is compared with the conventional methods. Section 3 also describes the ZIP model for assessing the treatment effect, and includes a comparison of the two-part statistic and the ZIP model using a simulation study. In Section 4, an application of the proposed method for calculating the sample size is illustrated in an example.

2. Two-part statistics and sample size

Consider a clinical trial to compare a test drug with a control. In the trial, \(N\) subjects are allocated to the treatment groups in the ratio \(r: 1-r\) \((0 < r < 1)\). A response variable for the primary endpoint \(Y_i\) follows the ZIP distribution that is given by

\[
P(Y_i) = \begin{cases} 
\omega_i + (1 - \omega_i)p_{C}(0) & (Y_i = 0) \\
(1 - \omega_i)p_{C}(Y_i) & (Y_i > 0)
\end{cases}
\]

\(p_{C}(Y_i) = \lambda_i^Y e^{-\lambda_i} / Y_i!\).

Here, \(\omega_i\) and \(\lambda_i\) are unknown parameters where \(0 < \omega_i < 1\) and \(\lambda_i > 0\). The case of the zero-deflated distribution \(\omega_i < 0\) is not covered in this article. The null hypothesis is

\(H_0: p_1 = p_2\) and \(\mu_1 = \mu_2\),

where \(p_i\) represents the proportion of zero counts

\[
p_i = P(Y_i = 0) = \omega_i + (1 - \omega_i)e^{-\lambda_i},
\]

and \(\mu_i\) represents the mean of non-zero counts \((i = 1, 2)\). In the non-zero part, \(Y_i\) follows the zero-truncated Poisson distribution and the mean \(\mu_i\) and the variance \(\sigma_i^2\) are given by

\[
\mu_i = E[Y_i \mid Y_i > 0] = \frac{\lambda_i}{1 - e^{-\lambda_i}},
\]

\[
\sigma_i^2 = V[Y_i \mid Y_i > 0] = \frac{\lambda_i}{1 - e^{-\lambda_i}} - \left(\frac{\lambda_i}{1 - e^{-\lambda_i}}\right)^2 e^{-\lambda_i}.
\]

The test statistic for the null hypothesis, called the two-part statistic, is defined as

\[X^2_{(2)} = X^2_{(1)} + U^2.\]

Here, \(X^2_{(1)}\) is the test statistic for the zero part and \(U^2\) is the test statistic for the non-zero part. The test statistic for the zero part \(X^2_{(1)}\) is

\[
X^2_{(1)} = \frac{(p_1 - p_2)^2}{\bar{p}(1 - \bar{p}) \left( \frac{1}{r} + \frac{1}{1 - r} \right)}.
\]
where $\bar{p} = rp_1 + (1 - r)p_2$. The test statistic for the non-zero part $U^2$ is derived from the square of the Wilcoxon test statistic or the $t$-test statistic. Assuming $X_{i1}^2$ and $U^2$ to be mutually independent, $X_{i1}^2$ follows the chi-square distribution with two degrees of freedom under the null hypothesis.

The sample sizes of the non-zero portions are

$$n'_1 = Nr(1 - p_1),$$
$$n'_2 = N(1 - r)(1 - p_2).$$

Given two independent random samples $Y_{1j}$ and $Y_{2k}$ ($j = 1, \ldots, n'_1; k = 1, \ldots, n'_2$), the Wilcoxon test statistic for the non-zero part $U$ is

$$U = \frac{\#(Y_1 > Y_2) + 0.5 \cdot \#(Y_1 = Y_2) - 0.5n'_1n'_2}{\sqrt{\frac{n'_1n'_2(n'_1 + n'_2 + 1)}{12} \left[ N'^3 - N' - \sum_t (t^3 - t) \right]}}$$

where $\#(Y_1 > Y_2)$ and $\#(Y_1 = Y_2)$ denote the number of pairs of $Y_{1j} > Y_{2k}$ and that of $Y_{1j} = Y_{2k}$, respectively, and $t$ denotes the number of ties for each pair of $Y_{1j} = Y_{2k}$.

When calculating the sample size, generally tie data are not taken into account due to the difficulty in the assumption for ties. However, if the response is an ordered categorical variable or $\lambda_i$ is small, there are considerable ties after rank transformation. In that case, the sample size may be underestimated. Zhao et al. (2008) suggested that the sample size of the Wilcoxon test be adjusted for ties. We perform this adjustment on the sample size for the two-part statistic using $U'$ instead of $U$ for calculating the power:

$$U' = \sqrt{\frac{A - 0.5}{1 - \sum_{c=1}^{D} (rP_t(c \mid \lambda_1) + (1 - r)P_t(c \mid \lambda_2))^3}}$$

where

$$A = \sum_{c=2}^{D} P_t(c \mid \lambda_1) \sum_{d=1}^{c-1} P_t(d \mid \lambda_2) + 0.5 \sum_{c=1}^{D} P_t(c \mid \lambda_1) P_t(c \mid \lambda_2).$$

Here, $A$ denotes $\Pr(Y_1 > Y_2) + 0.5 \Pr(Y_1 = Y_2)$, and $A$ becomes 0.5 under the null hypothesis. The denominator in the right side of equation (6) makes $U'$ follow the standard normal distribution. In equation (6), $P_t(y \mid \lambda_i)$ denotes the probability density of the zero-truncated distribution

$$P_t(Y = y \mid \lambda_i, Y > 0) = \frac{P(y)}{1 - P(0)} = \frac{\lambda_i^y}{y!(\lambda_i^y - 1)}.$$
sufficient to calculate the sample size, since the value of \( P_1(D > 20) \) is ignorable (<0.001) for the sample size estimation.

The power for the two-part statistic is given by

\[
Power = P_{\text{Chi}}(\chi^2_{\alpha,2df},2,X^2_{(2)})
\]  
(8)

where \( \chi^2_{\alpha,2df} \) denotes the \( \alpha \) percentage point of the chi-square distribution with two degrees of freedom. The notation \( P_{\text{Chi}} \) indicates the cumulative probability of the chi-square distribution at \( \chi^2_{\alpha,2df} \). The second parameter of two is the degrees of freedom. The third parameter of \( X^2_{(2)} \) is the non-centrality parameter, which can be calculated as

\[
X^2_{(2)} = X^2_{(1)} + U'^2
\]

\[
= N r (1 - r) \left\{ \frac{(p_1 - p_2)^2}{\bar{p}(1 - \bar{p})} + \frac{12p'(A - 0.5)^2}{1 - \sum_{c=1}^{D}(r P_t(c | \lambda_1) + (1 - r) P_t(c | \lambda_2))^3} \right\}
\]  
(9)

from equations (5) and (6), where \( p' = (1 - p_1)(1 - p_2)(1 - \bar{p})^{-1} \). At the planning stage, the sample size \( N \) is determined so that the power by equation (8) meets the requirements of the study.

Lachenbruch (2001a) also provided the two-part statistic using the \( t \)-test. When the two-part statistic uses the \( t \)-test statistic \( T \), it is given by

\[
X^2_{(2T)} = X^2_{(1)} + T^2
\]

\[
= N r (1 - r) \left\{ \frac{(p_1 - p_2)^2}{\bar{p}(1 - \bar{p})} + \frac{(\mu_1 - \mu_2)^2}{\sigma^2} p' \right\}
\]  
(10)

In clinical research, the \( t \)-test is sometimes used for data following the Poisson distribution. However, since the \( \lambda \) parameter of the ZIP distribution is generally not so large, the normal approximation of the Poisson distribution may not be accurate in this case. The \( t \)-test should thus be used with care for the two-part statistic.

3. Comparison

3.1 Comparison of two-part statistics using Wilcoxon test, Wilcoxon test adjusted for ties, and \( t \)-test

In Section 2, the power of two-part statistic using the Wilcoxon test adjusted for ties was introduced. We compared the power with and without the adjustment for ties. For calculating the power of the Wilcoxon test without the adjustment for ties, the following \( U'' \) (Noether, 1987) is used instead of \( U' \) in equation (9):

\[
U'' = \frac{(Pr(Y_1 > Y_2) - 0.5)n'_1n'_2}{\sqrt{n'_1n'_2(n'_1 + n'_2 + 1)}/12}
\]  
(11)
Furthermore, the power of the two-part statistic using the $t$-test based on equation (10) was compared with the power of the two-part statistic using the Wilcoxon test.

Figure 1 displays the power of the two-part statistic using the Wilcoxon test, the Wilcoxon test adjusted for ties, and the $t$-test. The power of the three methods is similar in most cases (Figure 1(a)). In some cases, the power of the two-part statistic using the $t$-test is the highest (Figure 1(b)). However, large differences in the power between the $t$-test and the other tests were not found for large $\lambda$. As mentioned previously, since the $\lambda$ parameter of the ZIP distribution is generally not so large, the $t$-test should be used with caution for the two-part statistic. The power of the two-part statistic using the Wilcoxon test is slightly higher than the power of that using the Wilcoxon test adjusted for ties. The power of the two-part statistic using the adjusted Wilcoxon test is very similar to the actual power based on a simulation study when the two-part statistic employs the Wilcoxon test. Therefore, the sample size can be reliably calculated using equation (6) whenever ties need to be taken into account.

3.2 Comparison with conventional tests

We compared the power of the two-part statistic with that of conventional methods ignoring zero-inflation in order to characterize the power of the two-part statistic. One of the conventional methods was the chi-square test postulating the null hypothesis of $H_{0C}: p_1 = p_2$. In this comparison, the power for the chi-square test was calculated using the normal approximation method. Another conventional method was the Wilcoxon test for the null hypothesis of $H_{0W}: \mu_{a1} = \mu_{a2}$, where $\mu_{ai}$ was the mean of all data following the ZIP distribution:

$$\mu_{ai} = (1 - \omega_i)\lambda_i.$$  \(\text{(12)}\)

The variance of the ZIP distribution is

$$\sigma_{ai}^2 = (1 - \omega_i)(1 + \omega_i\lambda_i).$$
In this case, the power of the Wilcoxon test without ties estimated by equation (11) is considerably different from the actual power. This difference in the power is much greater than that for the ordinary Poisson data without zero-inflation. The reason is that the ZIP data includes far more ties than the ordinary Poisson data due to many zero counts. Therefore, methods adjusted for ties should be applied to zero-inflated data whenever the sample size or power is calculated. In this article, the power was calculated using the method proposed by Zhao et al. (2008).

Figure 2 compares the two-part statistic using the Wilcoxon test adjusted for ties with the chi-square test and the Wilcoxon test when $N = 200$, $r = 0.5$, and $\alpha = 0.05$. When $\omega_1 = \omega_2$ and $\lambda_1 = \lambda_2$ ($\lambda_2 = 3$ in Figure 2(a) and $\omega_2 = 0.3$ in Figure 2(c)), the null hypotheses of $H_0$, $H_{0C}$, and $H_{0W}$ are true and the power is 0.05. In Figure 2(d), $H_{0C}$ is true for $\omega_2 = 0.322$ from equation (1), and $H_{0W}$ is true for $\omega_2 = 0.475$ from equation (12).

Given no difference in $\omega$, the two-part statistic had the highest power (Figure 2(a)). In the case of no difference in $\lambda$, the chi-square test gave the highest power (Figure 2(b)). Thus, the chi-square test was effective if the drug effect in the zero part was certain but the drug effect in the other part was not.

Fig. 2. Power of the two-part statistic and conventional methods of chi-square test and Wilcoxon test when $N = 200$, $r = 0.5$, and $\alpha = 0.05$. 
Fig. 3. Power of the two-part statistic and conventional methods of chi-square test and Wilcoxon test when $r = 0.5$, $\alpha = 0.05$, $\omega_1 = \omega_2 = 0.3$ and $\lambda_1 = 3$.

the non-zero part was uncertain, although the two-part statistic was not significantly inferior to the chi-square test. Given a difference in both of $\omega$ and $\lambda$, the Wilcoxon test or the two-part statistic had the highest power. Overall, the two-part statistic maintained a steady power.

Figure 3(a) shows the power of the two-part statistic for various values of $N$. The behavior of the power did not change depending on $N$ for different values of $\omega$ and $\lambda$. Figure 3(b) shows the power for the two-part statistic, the chi-square test, and the Wilcoxon test when $N$ is small. The relationship between the three methods was the same as that presented in Figure 2(a).

3.3 Comparison with the ZIP model

In this section, the ZIP model for assessing the treatment effect is described, and a comparison of the two-part statistic and the ZIP model is made using a simulation study. The ZIP model is a mixture model of the logit model for $\omega$ and the Poisson regression model for $\lambda$. The ZIP model is expressed by

$$
\text{logit}(\omega) = G\gamma,\\
\log(\lambda) = B\beta,
$$

where $B$ and $G$ represent covariate matrices. Lambert (1992) supposed that one could observe $Z_l = 1$ when $Y_l$ was from the inflated zero, and $Z_l = 0$ when $Y_l$ was from the Poisson state. The parameters $\gamma$ and $\beta$ are estimated by maximizing the log-likelihood:

$$
L(\gamma, \beta; y, z) = \sum_{l=1}^{N} \log(f(z_l | \gamma)) + \sum_{l=1}^{N} \log(f(y_l | z_l, \beta))
$$

$$
= \sum_{l=1}^{N} [z_l G_l \gamma - \log(1 + e^{G_l \gamma})] + \sum_{l=1}^{N} (1 - z_l) \log(y_l B_l \beta - e^{B_l \beta}) - \sum_{l=1}^{N} (1 - z_l) \log y_l!
$$

where $G_l$ and $B_l$ are the $l$-th rows of $G$ and $B$. On the right side of the equation, the first term including $\gamma$ and the second term including $\beta$ can be maximized separately. Recently, maximum
likelihood estimation procedures have been included in software packages such as the package pscl in R (Zeileis et al., 2008) and the COUNTREG procedure in SAS.

The treatment effect $T$ can be included in the covariate matrices $G$ and $B$. The null hypotheses for the treatment comparison are $H_{0\gamma}: \gamma_t = 0$ and $H_{0\beta}: \beta_t = 0$, where $\gamma_t$ and $\beta_t$ are the parameter estimates of the treatment effect. That is, there are two hypothesis tests for the treatment effect: one in zero-inflation and one in the mean of the Poisson distribution. On the other hand, the two-part statistic is one comprehensive test. In order to compare the power of testing the treatment effect in the ZIP model with the power of the two-part statistic, Fisher’s combination test (Fisher, 1948) is applied. Assume that the $p$-values $p_\gamma$ for $H_{0\gamma}$ and $p_\beta$ for $H_{0\beta}$ are mutually independent and uniformly distributed on $[0,1]$ under the null hypothesis. The level $\alpha$ combination test rejects the null hypotheses $H_{0\gamma}$ and $H_{0\beta}$ if

$$p_\gamma p_\beta < \exp\left(-\frac{1}{2}\chi^2_{1-\alpha,4}\right).$$

We compared power using a simulation study where the power was defined as rejected proportions of the null hypothesis on 10,000 replications. The estimates of the ZIP model were computed using the zeroinfl function of the package pscl in R. The ZIP model included the treatment effect alone. The two-part statistic was calculated using the chi-square test and the Wilcoxon test. The allocation ratio $r$ was 0.5, and $\alpha$ was set at 0.05.

Table 1 shows the simulation results for $N = 200$. The table contains the results for three cases: treatment difference in the zero-inflation $\omega$ alone ($\lambda_1 = \lambda_2 = 3$), treatment difference in the Poisson parameter $\lambda$ alone ($\omega_1 = \omega_2 = 0.3$), and treatment difference in both parameters ($\omega_1 = 0.3$, $\omega_2 = 0.1 - 0.5$, $\lambda_1 = 3$, and $\lambda_2 = 1 - 3$). When $N = 200$, the power of the combination test of the ZIP model and of the two-part statistic were generally similar. Differences were found between the test for zero-inflation of the ZIP model and the chi-square test component of the two-part statistic. The difference was large when $\omega_1$ and $\omega_2$ were equal or similar and $\lambda_2$ was 1 or 2. This was caused by the difference in the null hypotheses of $H_{0\gamma}: \gamma_t = 0$ and $H_0: p_1 = p_2$, where $p_1$ and $p_2$ included the zero counts from the Poisson state. If the treatment difference in the zero-inflation is of particular interest, then the ZIP model can detect this.

Table 2 shows the simulation results for $N = 60$. The two-part statistic had higher power than the combination test for the ZIP model in some cases (e.g., $\omega_1 = \omega_2 = 0.3$, $\lambda_1 = 3$, and $\lambda_2 = 1$; $\omega_1 = 0.3$, $\omega_2 = 0.4$, $\lambda_1 = 3$, and $\lambda_2 = 1$). This magnitude of difference was not found for $N = 200$, because the power for the two-part statistic and the combination test for the ZIP model was in excess of 0.99 in these cases. Regarding the combination test for the ZIP model, the power under the null hypothesis (i.e., $\omega_1 = \omega_2 = 0.3$, $\lambda_1 = \lambda_2 = 3$) was less than the nominal type I error rate of 0.05 although this was approximately 0.05 for $N = 200$. In the case of no treatment difference in the zero-inflation scenario (i.e., $\omega_1 = \omega_2$), the power for the zero-inflation of the ZIP model was also less than the nominal type I error rate of 0.05 in Table 2. In contrast, regarding the two-part statistics, the power under the null hypothesis was approximately 0.05
Table 1. Power of the zero-inflated Poisson model and the two-part statistic when a covariate was treatment effect alone, $N = 200$, $r = 0.5$, and $\alpha = 0.05$.

| $\omega_1$ | $\omega_2$ | $\lambda_1$ | $\lambda_2$ | Two-part | $\chi^2$ | Wilcoxon | Combination | Zero-inflation | Non-zero |
|------------|------------|-------------|-------------|----------|----------|----------|-------------|--------------|----------|
| 0.3        | 0.3        | 3           | 1           | >0.999   | 0.899    | >0.999   | >0.999      | 0.040        | >0.999   |
| 0.3        | 0.3        | 3           | 2           | 0.835    | 0.140    | 0.877    | 0.820       | 0.038        | 0.895    |
| 0.3        | 0.3        | 3           | 3           | 0.047    | 0.051    | 0.047    | 0.046       | 0.045        | 0.052    |
| 0.3        | 0.3        | 3           | 4           | 0.737    | 0.065    | 0.816    | 0.751       | 0.046        | 0.847    |
| 0.3        | 0.3        | 3           | 5           | 0.999    | 0.076    | >0.999   | >0.999      | 0.045        | >0.999   |
| 0.3        | 0.4        | 3           | 1           | >0.999   | 0.989    | >0.999   | >0.999      | 0.158        | >0.999   |
| 0.3        | 0.4        | 3           | 2           | 0.926    | 0.564    | 0.851    | 0.894       | 0.243        | 0.878    |
| 0.3        | 0.4        | 3           | 3           | 0.214    | 0.281    | 0.048    | 0.199       | 0.272        | 0.048    |
| 0.3        | 0.4        | 3           | 4           | 0.760    | 0.196    | 0.786    | 0.806       | 0.283        | 0.810    |
| 0.3        | 0.4        | 3           | 5           | >0.999   | 0.166    | 0.999    | 0.999       | 0.291        | >0.999   |
| 0.3        | 0.1        | 3           | 3           | 0.830    | 0.898    | 0.048    | 0.685       | 0.838        | 0.049    |
| 0.3        | 0.2        | 3           | 3           | 0.234    | 0.315    | 0.043    | 0.204       | 0.289        | 0.050    |
| 0.3        | 0.3        | 3           | 3           | 0.047    | 0.051    | 0.047    | 0.046       | 0.045        | 0.052    |
| 0.3        | 0.4        | 3           | 3           | 0.213    | 0.281    | 0.048    | 0.199       | 0.272        | 0.048    |
| 0.3        | 0.5        | 3           | 3           | 0.674    | 0.778    | 0.049    | 0.654       | 0.761        | 0.049    |
| 0.3        | 0.1        | 3           | 4           | 0.995    | 0.969    | 0.862    | 0.992       | 0.907        | 0.885    |
| 0.3        | 0.2        | 3           | 4           | 0.894    | 0.474    | 0.846    | 0.889       | 0.322        | 0.862    |
| 0.3        | 0.3        | 3           | 4           | 0.737    | 0.065    | 0.816    | 0.751       | 0.046        | 0.847    |
| 0.3        | 0.4        | 3           | 4           | 0.760    | 0.196    | 0.786    | 0.806       | 0.283        | 0.810    |
| 0.3        | 0.5        | 3           | 4           | 0.914    | 0.710    | 0.748    | 0.936       | 0.784        | 0.774    |

and did not change for $N = 200$.

4. Example

We consider a clinical trial whose aim is to compare a hormone therapy to a placebo for menopausal women with vasomotor symptoms. Hot flushes are common in menopausal women (CHMP, 2005). The daily frequency of moderate to severe hot flushes is the primary endpoint of the trial. When the trial subjects are patients with mild symptoms, in the range of three to twelve hot flushes daily, hot flushes can be completely removed in a certain proportion of subjects by hormone therapy and even by placebo treatment.

Assume that the theoretical data in Figure 4 have been obtained in a previous study. The ZIP parameters estimated from the data are $\omega_1 = 0.42$ and $\lambda_1 = 3.77$ for the hormone therapy and $\omega_2 = 0.21$ and $\lambda_2 = 4.57$ for the placebo using the zeroinfl function in R or the COUNTREG procedure in SAS.

If individual data are not available but the mean of non-zero counts $\mu_i$ and the proportion
Table 2. Power of the zero-inflated Poisson model and the two-part statistic when a covariate was treatment effect alone, $N = 60$, $r = 0.5$, and $\alpha = 0.05$.

| $\omega_1$ | $\omega_2$ | $\lambda_1$ | $\lambda_2$ | Power for Two-partStatistic | Power for ZIP-model |
|------------|-------------|-------------|-------------|-----------------------------|---------------------|
|            |             |             |             | Two-part | $\chi^2$ | Wilcoxon | Combination | Zero-inflation | Non-zero |
| 0.3        | 0.3         | 3           | 1           | 0.932    | 0.417   | 0.910    | 0.857   | 0.032   | 0.929 |
| 0.3        | 0.3         | 3           | 2           | 0.316    | 0.071   | 0.382    | 0.278   | 0.020   | 0.412 |
| 0.3        | 0.3         | 3           | 3           | 0.047    | 0.046   | 0.046    | 0.032   | 0.027   | 0.044 |
| 0.3        | 0.3         | 3           | 4           | 0.250    | 0.049   | 0.325    | 0.255   | 0.032   | 0.357 |
| 0.3        | 0.3         | 3           | 5           | 0.756    | 0.054   | 0.826    | 0.773   | 0.037   | 0.863 |
| 0.3        | 0.4         | 3           | 1           | 0.948    | 0.611   | 0.874    | 0.837   | 0.079   | 0.888 |
| 0.3        | 0.4         | 3           | 2           | 0.390    | 0.206   | 0.344    | 0.291   | 0.082   | 0.368 |
| 0.3        | 0.4         | 3           | 3           | 0.094    | 0.117   | 0.045    | 0.070   | 0.093   | 0.042 |
| 0.3        | 0.4         | 3           | 4           | 0.275    | 0.093   | 0.315    | 0.307   | 0.090   | 0.345 |
| 0.3        | 0.4         | 3           | 5           | 0.720    | 0.087   | 0.801    | 0.777   | 0.097   | 0.849 |
| 0.3        | 0.1         | 3           | 3           | 0.320    | 0.408   | 0.046    | 0.065   | 0.941   | 0.044 |
| 0.3        | 0.2         | 3           | 3           | 0.098    | 0.125   | 0.045    | 0.047   | 0.052   | 0.046 |
| 0.3        | 0.3         | 3           | 3           | 0.047    | 0.046   | 0.046    | 0.032   | 0.027   | 0.044 |
| 0.3        | 0.4         | 3           | 3           | 0.094    | 0.117   | 0.045    | 0.070   | 0.093   | 0.042 |
| 0.3        | 0.5         | 3           | 3           | 0.245    | 0.320   | 0.046    | 0.191   | 0.280   | 0.046 |
| 0.3        | 0.1         | 3           | 4           | 0.649    | 0.535   | 0.363    | 0.476   | 0.234   | 0.377 |
| 0.3        | 0.2         | 3           | 4           | 0.372    | 0.176   | 0.344    | 0.333   | 0.095   | 0.371 |
| 0.3        | 0.3         | 3           | 4           | 0.250    | 0.049   | 0.325    | 0.255   | 0.032   | 0.357 |
| 0.3        | 0.4         | 3           | 4           | 0.275    | 0.093   | 0.315    | 0.307   | 0.090   | 0.345 |
| 0.3        | 0.5         | 3           | 4           | 0.405    | 0.278   | 0.283    | 0.442   | 0.286   | 0.314 |

Fig. 4. Daily frequency of moderate to severe hot flushes.
of zero counts $p_i$ are known, the ZIP parameters can be roughly estimated as follows. Suppose that we know only $\mu_1 = 3.86$ and $p_1 = 0.43$ for the hormone therapy, and $\mu_2 = 4.62$ and $p_2 = 0.22$ for the placebo. The mean $\mu_i$ is equal to $\lambda_i/[1 - \exp(-\lambda_i)]$ by equation (3). Using the iteration scheme, $\lambda_i^{(t+1)} = \mu_i [1 - \exp(\lambda_i^{(t)})]$, where $\lambda_i^{(t)}$ is the $t$-th estimate (Dietz et al., 2000), $\lambda_1$ converges to 3.77 for the hormone therapy and $\lambda_2$ to 4.57 for the placebo. Then, $\omega_1 = 0.42$ for the hormone therapy and $\omega_2 = 0.20$ for the placebo are estimated from $\omega_i = [p_i - \exp(-\lambda_i)]/[1 - \exp(-\lambda_i)]$ based on equation (2).

The sample size for the two-part statistic is calculated from equations (8) and (9). In equation (6), $A = 0.601$. For a 90% power, the sample size adjusting for ties is 158 when $r = 0.5$ and $\alpha = 0.05$.

5. Discussion

In clinical trials, it is often undesirable to perform two hypothesis tests in the primary analysis because of the difficulty in interpreting the two possibly inconsistent test results and controlling the type I error inflation. The two-part statistic is effective in terms of producing one test result as well as embracing the treatment effect in the zero and the non-zero parts. We provided the methods for calculating sample size and power for the two-part statistic using the Wilcoxon test adjusted for ties. The power estimated by our method was very similar to the actual power based on a simulation study when the two-part statistic employed the Wilcoxon test. In our study, the two-part statistic showed higher power than conventional tests in most cases. Even when a conventional test showed higher power than the two-part statistic, the difference was small. However, if the complete recovery from a disease is of primary interest in the clinical research, the chi-square test on a binary response of zero vs. non-zero outcome should be used for the primary analysis.

The ZIP model can estimate the profile of the ZIP distribution and the extent of zero-inflation. For example, Cheung (2002) applied the ZIP model to child growth and motor development and explored the effect of covariates. However, our study focused on the treatment effect in the ZIP model for comparison with the two-part statistic. The simulation results showed that the power of the combination test for the ZIP model and the two-part statistic were generally similar when the covariate was the treatment effect alone. In some results with a small sample size, the two-part statistic demonstrated higher power than the ZIP model. The ZIP model is suitable for characterizing the distribution and exploring the effect of some covariates but may be unsuitable for aiming at the detection of treatment difference especially when the sample size is not large. In our simulation study, we only considered the treatment effect in the ZIP model for the purpose of comparison with the two-part statistic. However, in practice, the ZIP model can include some covariates in the logit model for $\omega$ and the Poisson regression model for $\lambda$.

The two-part statistic can be adjusted with covariates. For example, a logistic regression
model is used for the zero part. For the non-zero part, ANCOVA with ranks transformed from $Y$ can be applied (Conover et al., 1982). It is a subject for further research to develop methods for the sample size estimation in this context and to assess the effect of covariates on the power.

Acknowledgements

The authors are grateful to an anonymous referee for insightful suggestions on this manuscript.

REFERENCES

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.

Committee for Medical Products for Human Use (CHMP) (2005). Guideline on clinical investigation of medical products for hormone replacement therapy of oestrogen deficiency symptoms in postmenopausal women. European Medicines Agency: London.

Conover W. J. and Imran R. L. (1982). Analysis of covariance using the rank transformation. *Biometrics* 38, 715–724.

Delucchi K. L., Boström A. (2004). Methods for analysis of skewed data distributions in psychiatric clinical studies: Working with many zero values. *Am J Psychiatry* 161, 1159–1168.

Dietz E. and Böhning D. (2000). On estimation of the Poisson parameter in zero-inflated Poisson models. *Computational Statistics and Data Analysis* 34, 441–459.

Fisher R. A. (1948). Questions and answers: Combining independent tests of significance. *The American Statistician* 2, 30.

Hallstorm A. P. (2010). A modified Wilcoxon test for non-negative distributions with a clump of zeros. *Statistics in Medicine* 29, 391–440.

Heilbron D. C. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal* 36, 531–547.

Lachenbruch P. A. (2001a). Comparisons of two-part models with competitors. *Statistics in Medicine* 20, 1215–1234.

Lachenbruch P. A. (2001b). Power and sample size requirements for two-part models. *Statistics in Medicine* 20, 1235–1238.

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

Noether G. E. (1987). Sample size determination for some common nonparametric tests. *Journal of the American Statistical Association* 82(398), 645–647.
Zeileis A., Kleiber C., Jackman S. (2008). Regression models for count data in R. *Journal of Statistical Software* **27**(8).

Zhao Y. D., Rahardja D., Qu1 Y. (2008). Sample size calculation for the Wilcoxon-Mann-Whitney test adjusting for ties. *Statistics in Medicine* **27**, 462–468.