Journal of Biopharmaceutical Statistics

ISSN: 1054-3406 (Print) 1520-5711 (Online) Journal homepage: https://www.tandfonline.com/loi/lbps20

On the Use of Zero-Inflated and Hurdle Models for Modeling Vaccine Adverse Event Count Data

C. E. Rose , S. W. Martin , K. A. Wannemuehler & B. D. Plikaytis

To cite this article: C. E. Rose , S. W. Martin , K. A. Wannemuehler & B. D. Plikaytis (2006) On the Use of Zero-Inflated and Hurdle Models for Modeling Vaccine Adverse Event Count Data, Journal of Biopharmaceutical Statistics, 16:4, 463-481, DOI: 10.1080/10543400600719384

To link to this article: https://doi.org/10.1080/10543400600719384

Journal of Biopharmaceutical Statistics, 16: 463–481, 2006

ISSN: 1054-3406 print/1520-5711 online
DOI: 10.1080/10543400600719384

Taylor & Francis
Taylor & Francis Group

ON THE USE OF ZERO-INFLATED AND HURDLE MODELS FOR MODELING VACCINE ADVERSE EVENT COUNT DATA

C. E. Rose and S. W. Martin
Bacterial Vaccine-Preventable Diseases Branch, Division of Epidemiology and Surveillance, Atlanta, Georgia, USA

K. A. Wannemuehler and B. D. Plikaytis
Biostatistics Office, Division of Bacterial and Mycotic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

We compared several modeling strategies for vaccine adverse event count data in which the data are characterized by excess zeroes and heteroskedasticity. Count data are routinely modeled using Poisson and Negative Binomial (NB) regression but zero-inflated and hurdle models may be advantageous in this setting. Here we compared the fit of the Poisson, Negative Binomial (NB), zero-inflated Poisson (ZIP), zero-inflated Negative Binomial (ZINB), Poisson Hurdle (PH), and Negative Binomial Hurdle (NBH) models. In general, for public health studies, we may conceptualize zero-inflated models as allowing zeroes to arise from at-risk and not-at-risk populations. In contrast, hurdle models may be conceptualized as having zeroes only from an at-risk population. Our results illustrate, for our data, that the ZINB and NBH models are preferred but these models are indistinguishable with respect to fit. Choosing between the zero-inflated and hurdle modeling framework, assuming Poisson and NB models are inadequate because of excess zeroes, should generally be based on the study design and purpose. If the study's purpose is inference then modeling framework should be considered. For example, if the study design leads to count endpoints with both structural and sample zeroes then generally the zero-inflated modeling framework is more appropriate, while in contrast, if the endpoint of interest, by design, only exhibits sample zeroes (e.g., at-risk participants) then the hurdle model framework is generally preferred. Conversely, if the study's primary purpose it is to develop a prediction model then both the zero-inflated and hurdle modeling frameworks should be adequate.

Key Words: Excess zeroes; Hurdle model; Negative binomial; Poisson; Vaccine adverse events; Zero-inflated model.

1. INTRODUCTION

We compare and contrast several modeling strategies for overdispersed vaccine adverse event count data. Data from public health studies often include count

Received and Accepted March 13, 2006
Address correspondence to C. E. Rose, Bacterial Vaccine-Preventable Diseases Branch, Division of Epidemiology and Surveillance, CDC Mailstop C-09, 1600 Clifton Road NE, Atlanta, GA 30333, USA; Fax: 404-639-3266; E-mail: cvr7@cdc.gov

endpoints which frequently exhibit excess zeroes and/or heteroskedasticity. These data are routinely modeled using Poisson and negative binomial (NB) regression. However, zero-inflated and hurdle models, which have gained in popularity in recent years for modeling count data, may be advantageous in this setting. Our motivating data is from an anthrax vaccine absorbed (AVA) clinical trial study, where the number of systemic adverse events occurring after each of four injections was collected for each participant. We assessed the model fit of Poisson, Negative Binomial (NB), zero-inflated Poisson (ZIP), zero-inflated Negative Binomial (ZINB), Poisson Hurdle (PH) and Negative Binomial Hurdle (NBH) models. We utilize the robust variance-covariance estimator suggested by Moulton et al. (2002) to account for the repeated measurements on participants.

1.1. Background

Poisson and NB regression are typically used in epidemiology to model count data. Over-dispersion in count data occurs frequently due to excess zeroes, unexplained heterogeneity, and/or temporal dependency (Cameron and Trivedi, 1998). Excess zeroes can be categorized as either sampling or structural zeroes. For example, in response to the question "How often did you drink alcohol during the last 30 days?" there will be individuals who drink alcohol but chose not to drink during the last 30 days (sampling zeroes) and individuals who never drink alcohol (structural zeroes).

Poisson regression is commonly used in public health to model the number of events (y) or rate (r); however the Poisson model assumes the variance is equal to the mean, an assumption that is often violated. The NB model has a built-in dispersion parameter that can account for variance greater than the mean (over-dispersion) that is due to unobserved heterogeneity and/or temporal dependency (Chin and Quddus, 2003). Zero-inflated and hurdle models are generally used in the setting of excess zeroes. Zero-inflated models are typically used if the data contains excess structural and sampling zeroes, whereas hurdle models are generally used when there are only excess sampling zeroes.

Fitted zero-inflated and hurdle models are often indistinguishable using goodness of fit statistics (Gray, 2005) but one model type may be more applicable based on the study objectives. Cheung (2002) suggested that the zero-inflated model is useful if there are two underlying disease processes, one which puts the subject at risk and the other which influences the outcome in the at-risk population. For example, Sheu et al. (2004) used the zero-inflated model that treated nonsmokers as structural zeroes, which clearly distinguishes them from the smokers. In contrast, if all individuals are considered at risk of an event then the realization of the event represents a hurdle that has been crossed, and the hurdle model may be more appropriate.

Zero-inflated models have been extensively used in other fields such as econometrics since Lambert's (1992) publication outlining the zero-inflated Poisson (ZIP) model. Recent medical applications of zero-inflated models are illustrated by Cheung (2002) who applied zero-inflated models to early growth and motor development and Bohning et al. (1999) who found the ZIP model useful for modeling the number of decayed, missing, or filled teeth. Moulton et al. (2002)

used a zero-inflated Bernoulli/log-gamma model to model HIV-1 RNA viral load assays. Hurdle models suggest a two-part process and have become more common for modeling health care utilization data. The first part induces an event, and once the hurdle to the first event has been cleared, the second part determines the number of subsequent events. Thus, the hurdle model conceptual framework is different from the zero-inflated models, which assume an initial process brings individuals into the at-risk population. For example, generally the patient initiates contact with a provider and once this hurdle is crossed the provider typically influences the number of future health care encounters. A Mexico health care utilization study used a hurdle model (Brown et al., 2005) (a zero-truncated NB and probit model) to model health care demand. Similarly, a study of individuals' health insurance purchasing decisions used a hurdle model (logit and OLS) for analyzing health insurance demand (Liu and Chen, 2002). In addition, Krobot et al. (2005) extended the concept of the hurdle model to include person-time.

Extensions of zero-inflated and hurdle regression models for correlated data have recently begun to appear for subject-specific and marginal models. Hall (2000), Yau and Lee (2001), and Berk and Lachenbruch (2002) included cluster-specific random effects into their zero-inflated models. Min and Agresti (2005) fit several random effect hurdle models to pharmaceutical data. They estimated the number of episodes of side effects that were recorded at each visit and compared two treatments. Several studies have fit marginal models for clustered data with excess zeroes. Dobbie and Welsh (2001) and Hall and Zhang (2004) used a generalized estimating equations (GEEs) or an extension thereof to estimate hurdle and zero-inflated marginal models, respectively.

Here we compare and contrast several modeling strategies using data from clinical trial participants. Our endpoint of interest is the number of unique systemic adverse events reported after each injection for each of our study participants. By design, all study participants are at risk of experiencing one or more systemic adverse events after a given injection, therefore all zeroes arise from sampling.

2. MOTIVATING EXAMPLE

Models discussed in this paper were fit using data from a multicenter, double-blind, randomized, placebo-controlled human AVA clinical trial. The study was designed to assess the immunogenicity and safety of AVA administered in full and reduced-dose regimens, and/or by an alternate route of administration. A total of 1564 subjects (aged 18–61 years) were enrolled in the study by investigators at five U.S. medical centers. The study was approved by the institutional review committees at each study center and CDC. Study centers include Walter Reed Army Institute of Research, Baylor College of Medicine, Emory University School of Medicine, Mayo Clinic and Foundation, and University of Alabama at Birmingham.

Here, for modeling purposes, we use the imputed, intent-to-treat, systemic adverse event data from the interim analysis. The interim analysis was conducted using the first 7 months of data collected from the first 1005 participants and includes the first 4 study injections. Study injections are scheduled for time 0, 2 and 4 weeks, and 6 months. Participants by gender are approximately equal, and this distribution remains consistent within each treatment group. The distribution

of participants by race is black (19%), white (76%), and other (5%), which remains consistent across the treatment groups. The distribution of participants by study sites remains consistent across the treatment groups.

A complete study description and study interim results will be presented in a later publication. The data used for this paper are intended to demonstrate the modeling techniques, and are not meant to provide an overview of the interim analysis results. The interim analysis treatment groups are:

Group I: Currently licensed regimen given subcutaneously (SQ)
Group II: Current full-dose regimen given intramuscularly (IM)
Group III: Reduced dosing regimen given IM
Group IV: Placebo (saline) given IM
Group V: Placebo (saline) given SQ

While missing reactogenicity outcomes were multiply imputed (to be described elsewhere), for simplicity of presentation we used only the first imputation for our model descriptions. We modeled the number of unique systemic adverse events occurring after an injection for each participant. Systemic adverse events included both solicited and unsolicited events. Solicited events included fatigue, muscle ache, headache, temperature, and tender/painful axilary adenopathy. Unsolicited events included all other adverse events that are not classified as injection site reactions or injuries.

Primary safety objectives were to assess the impact of change of route of vaccine administration on reactogenicity for all participants and to assess differences in reactogenicity between males and females. Our analysis of the total count of unique systemic adverse events will compare the licensed regimen Group I to the Group II regimen.

3. METHODS

We considered six models for modeling the number of systemic events occurring after a dose for a participant. These models are Poisson and NB, which we refer to as the standard models, zero-inflated Poisson (ZIP), zero-inflated NB (ZINB), Poisson hurdle (PH), and NB hurdle (NBH) models. We use the logit model with a linear predictor of $\eta = \beta' x$ for the zero-inflated and hurdle portions of all zero-inflated and hurdle models.

3.1. Poisson

Poisson regression is commonly used in public health to model the number of events (y) or rate (r). It's often assumed that the number of events follows a Poisson distribution with a conditional mean (μ) depending upon a set of regressors ($\boldsymbol{x}$) and corresponding parameters ($\boldsymbol{\beta}$) for a participant's linear predictor. Using a log link we can express the expected number of events for participant i at dose j as $\mu_{ij} = E(y_{ij} \mid \boldsymbol{x}_{ij}) = e^{\beta' \boldsymbol{x}_{ij}}$. The Poisson probability distribution of y_{ij} given $\boldsymbol{x}_{ij}$ can be expressed as

$$P(Y_{ij} = y_{ij}) = \frac{e^{-\mu_{ij}} \mu_{ij}^{y_{ij}}}{y_{ij}!} \tag{1}$$

where y is a non-negative integer. The contribution of the ith participant at dose j to the log-likelihood for the Poisson model can be expressed as

$$LL(\boldsymbol{\beta}) = y_{ij}(\boldsymbol{\beta}\boldsymbol{x}_{ij}) - e^{\boldsymbol{\beta}\boldsymbol{x}_{ij}} - \ln(y_{ij}!)$$

The log-likelihood of all other considered models can be expressed using their respective distribution. A property of Poisson regression, one often violated, is that the variance equals the mean, i.e., $\mathrm{Var}(y) = \sigma^2 = \mu$. If over-dispersion is an issue (i.e., variance exceeds the mean) the estimated parameters based on Poisson regression will be inefficient (Cameron and Trivedi, 1998). Generally, over-disperion is caused by unobserved heterogeneity, temporal dependency, and/or excess zeroes in the data. Unobservable heterogeneity is likely to be an issue since the AVA trial enrolled participants with significant variation in their socioeconomic and health-related factors. Temporal dependency due to multiple injections over time for each participant may be an issue. In addition, there may be excess zeroes because it's expected that many participants will not experience any systemic adverse events during the time periods monitored.

3.2. Negative Binomial

We can relax the variance assumption of Poisson regression and allow for an over-dispersion parameter by using the Negative Binomial model. Over-dispersion may be accounted for when using the NB model because of the addition of an error term, e, to the conditional mean of the Poisson regression model (Sheu et al., 2004), i.e., $\mu_{ij} = e^{(\boldsymbol{\beta}'\boldsymbol{x}_{ij} + e_{ij})}$. We normally assume that $\exp(e_{ij})$ has a gamma distribution with mean 1 and variance a so that the conditional mean of y_{ij} is still μ_{ij} but the conditional variance of y_{ij} becomes $\mu_{ij}(1 + a\mu_{ij})$. As a approaches zero, y becomes a Poisson distribution and as a becomes larger the distribution becomes more dispersed. The NB probability distribution for participant i at dose j is given by:

$$P(Y_{ij} = y_{ij}) = \frac{\boldsymbol{\Gamma}(y_{ij} + 1/a)}{\boldsymbol{\Gamma}(y_{ij} + 1)\boldsymbol{\Gamma}(1/a)} \frac{(a\mu_{ij})^{y_{ij}}}{(1 + a\mu_{ij})^{y_{ij}+1/a}} \tag{2}$$

Where μ_{ij}, a, and $\Gamma(\bullet)$ refer to the mean of the count distribution, the NB dispersion parameter, and the gamma function. The NB model is generally adequate for addressing over-dispersion due to unobserved heterogeneity and/or temporal dependency, but may be inadequate for over-dispersion resulting from excess zeroes.

3.3. Zero-Inflated Models

In recent years, zero-inflated and hurdle models have gained popularity for modeling count data with excess zeroes. According to Cameron and Trivedi (1998), zero-inflated and hurdle models can be viewed as finite mixture models with a degenerate distribution whose mass is concentrated at zero. Excess zeroes arise when the event of interest is not experienced by many of the subjects.

3.3.1. Zero-Inflated Poisson. The zero-inflated Poisson distribution for participant i at dose j can be defined as

$$P(Y_{ij} = y_{ij}) = \begin{cases} p_{ij} + (1 - p_{ij})e^{-\mu_{ij}} & y_{ij} = 0 \\ (1 - p_{ij})\dfrac{e^{-\mu_{ij}}\mu_{ij}^{y_{ij}}}{y_{ij}!} & y_{ij} > 0 \end{cases} \tag{3}$$

The probability of being an excess zero (p_{ij}) in Eq. (3) is often modeled using logistic regression. Here, for all zero-inflated and hurdle models, we use the logistic model to estimate p_{ij}. Hence, p_{ij} is estimated using

$$p_{ij} = \frac{1}{1 + e^{-\eta_{ij}}} \tag{4}$$

where η_{ij} is related to a set of explanatory variables ($\boldsymbol{x}$). Zero-inflated models put more weight on the probability of observing a zero by using a mixing distribution. Hence, for ZIP model (3) the probability of observing a zero is given by the sum of observing an excess zero plus the probability of observing a zero in the Poisson model. As illustrated, the ZIP model allows for two separate processes. Conceptually, the first step models the structural zeroes (e.g., logistic regression) and the second step models the Poisson distribution conditional on the excess zeroes, i.e., Poisson regression models the sampling zeroes and counts. The mean and variance of the ZIP model are given by

$$E(y) = (1 - p)\mu$$
$$\mathrm{Var}(y) = \mu(1 - p)(1 + \mu p)$$

It can be seen from the ZIP mean and variance that when p equals zero the ZIP model reduces to the standard Poisson model. In contrast, as p approaches one the variance increases and the data exhibit greater overdisperion. The over-dispersion accounted for in the ZIP model is conceptually a result of the structural zeroes. Interpretation of the ZIP model depends upon what is being modeled. For medical studies the zero-inflated portion can be thought of as the odds of moving from the non-risk to the at-risk group. Once in the at-risk group we can determine the expected number of events or the risk of an event for one group versus another group.

3.3.2. Zero-Inflated Negative Binomial. Zero-inflated Negative Binomial (ZINB) models are sometimes preferred because they allow for additional flexibility in the variance. Using Eq. (2) we can express the ZINB model for participant i at dose j as

$$P(Y_{ij} = y_{ij}) = \begin{cases} p_{ij} + (1 - p_{ij})\dfrac{1}{(1 + a\mu_{ij})^{1/a}} & y_{ij} = 0 \\ (1 - p_{ij})\dfrac{\Gamma(y_{ij} + 1/a)}{\Gamma(y_{ij} + 1)\Gamma(1/a)}\dfrac{(a\mu_{ij})^{y_{ij}}}{(1 + a\mu_{ij})^{y_{ij}+1/a}} & y_{ij} > 0 \end{cases} \tag{5}$$

where all terms have been defined previously and the mean is as for the ZIP model but the variance is given by $\sigma^2 = \mu(1-p)[1+\mu(p+\alpha)]$. Note that the variance depends on p and the dispersion parameter α. The ZINB model allows for added flexibility compared to the ZIP model. It allows for over-dispersion arising from excess zeroes and heterogeneity, whereas the ZIP model only accommodates over-dispersion from excess zeroes. Interpretation of the ZINB model is as for the ZIP model.

3.4. Hurdle Models: Poisson and Negative Binomial

In contrast to zero-inflated models, hurdle models can be interpreted as two-part models. The first part is typically a binary response model and the second part is usually a truncated-at-zero count model (Cameron and Trivedi, 1998). Hence, the hurdle model is a modified count model in which separate processes generating the zeroes and positive counts are not constrained to be the same. This allows us to interpret the positive outcomes (>0) that result from passing the zero hurdle (threshold). The hurdle portion of the two-part model estimates the probability that the threshold is crossed. Theoretically the threshold could be any value, but it's usually taken as zero because this is most often meaningful in the context of the study objectives. Mullahy (1986) laid out the basic form of hurdle count models. Assume that f_1 and f_2 are any probability density functions for non-negative integers. A hurdle model can be expressed as

$$\begin{aligned} P[y=0] &= f_1(0) = p \\ P[y=i] &= (1-p)\frac{f_2(i)}{1-f_2(0)} = (1-p)f_2'(i) \quad i>0 \end{aligned} \tag{6}$$

Note that $f_1(\bullet)$ governs the hurdle part and $f_2(\bullet)$ the count process once the hurdle has been crossed. Furthermore, $f_1(0)$ is the probability of crossing the hurdle and $f_2'(\bullet)$ is the truncated normalization of $f_2(\bullet)$. Note that if $f_1(\bullet) = f_2(\bullet)$ the hurdle model collapses to the standard count model. Hurdle models can be specified in various ways by choosing different distributions for $f_1(\bullet)$ and $f_2(\bullet)$. As for the zero-inflated models we use logistic regression to model p. Here we define two hurdle models by specifying $f_2(\bullet)$ as the Poisson and NB distributions. For example, substitution of Eq. (1) into Eq. (6) results in the Poisson hurdle model for participant i at dose j being defined as

$$P(Y_{ij}=y_{ij}) = \begin{cases} p_{ij} & y_{ij}=0 \\ (1-p_{ij})\dfrac{e^{-\mu_{ij}}\mu_{ij}^{y_{ij}}}{(1-e^{-\mu_{ij}})y_{ij}!} & y_{ij}>0 \end{cases} \tag{7}$$

All terms are as defined previously and specification of the log-likelihood can be obtained using Eq. (7). The expected value for the Poisson hurdle (PH) model is given by

$$E(y) = \frac{(1-p)\mu}{1-f_2(0)}.$$

Substitution of Eq. (2) into Eq. (6) for $f_2(\bullet)$ results in the Negative Binomial hurdle (NBH) model. Computing the expected value for the NBH model is as for the PH model.

3.5. Parameter Robust Variance-Covariance Estimation

We expect some serial correlation because of repeated measurements for each participant, i.e., number of systemic adverse event data are collected for the ith participant at dose j ($j = 1, 2, 3, 4$). Here we use the strategy of Moulton et al. (2002) and first estimate the zero-inflated and hurdle models by treating the data as independent during the fitting process and then estimate a robust variance-covariance matrix. Specifically, we use the robust variance-covariance estimator suggested by Moulton et al. (2002) of

$$V\left[\sum_{i=1}^{N}\sum_{j=1}^{4} s_{ij}s'_{ij}\right]V$$

where V is the standard variance-covariance estimator and s_{ij} is the ith participant's contribution to the score vector at dose j.

4. MODEL SELECTION

Selecting the "best" model is often based on quantitative results and qualitative information, i.e., selecting the "best" model is often as much art as science. Our qualitative assessment of considered models is based on model appropriateness. Here appropriateness is defined in terms of model assumptions that can be assessed qualitatively such as interpretation of model parameters. Since our response variable, total number of systemic adverse events recorded for a participant after an injection, is a non-negative integer we assume Poisson and NB models are appropriate. Parameters from both models can be interpreted as a risk ratio, which is desired since our study focuses on assessing if one group has an increased risk compared to another group for the occurrence of systemic adverse events. We assume all participants are at risk of having one or more systemic adverse event(s) (e.g., observed zeroes arise from sampling only); therefore, we deem the hurdle model more appropriate than the zero-inflated model. However, this does not imply that the Poisson/NB hurdle models are more appropriate than the standard Poisson/NB models. Here our purpose is to compare and contrast these models; hence, we fit and compare zero-inflated models with hurdle and standard Poisson/NB models.

We used several criteria to compare and select among considered models. Nested models were tested using a likelihood ratio and/or score test. The Poisson model is nested within the NB model and as a result we can test if there is over-dispersion due to heterogeneity by testing if the dispersion parameter is necessary. We tested if the dispersion parameter was necessary by comparing the Poisson versus NB, ZIP versus ZINB, and PH versus NBH models. In addition, the Poison and NB models are nested within the ZIP and ZINB models, respectively, and these models were tested for over-dispersion due to excess zeroes. The Poisson and

ZIP, NB, and ZINB are not nested within the PH and NBH models, respectively. Therefore, we used the test proposed by Vuong (1989) for non-nested models to distinguish among these models. The Vuong statistic is given by

$$V = \frac{\bar{m}\sqrt{n}}{S_m}$$

where $m_i = \ln\left[\frac{\hat{P}_1(Y_i|X_i)}{\hat{P}_S(Y_i|X_i)}\right]$, P_S is usually taken to be the standard Poisson/NB and P_1 is the zero-inflated/hurdle model. The statistic m_i has a mean $\bar{m}$ and standard deviation S_m. The statistic V asymptotically follows a standard normal distribution. If $V > 1.96$ then it favors the zero-inflated model and $V < -1.96$ favors the standard model. In addition we assessed the general fit of the models by computing Pearson's chi-square using the predicted counts. Lastly, we used the Akaike (AIC) and Bayesian (BIC) information criteria to compare models.

5. RESULTS

We computed the empirical mean and variance for our systemic adverse event endpoint. Total number of systemic adverse events was recorded after each of the four injections for the 1005 study participants, which results in 4020 observations. Study participants experienced from 0–12 unique systemic adverse events after an injection. The observed mean and variance using all 4020 observations are 1.51 and 2.90, respectively. Our observed variance to mean ratio is 1.92, which indicates some over-dispersion. Figure 1 compares the observed count distribution

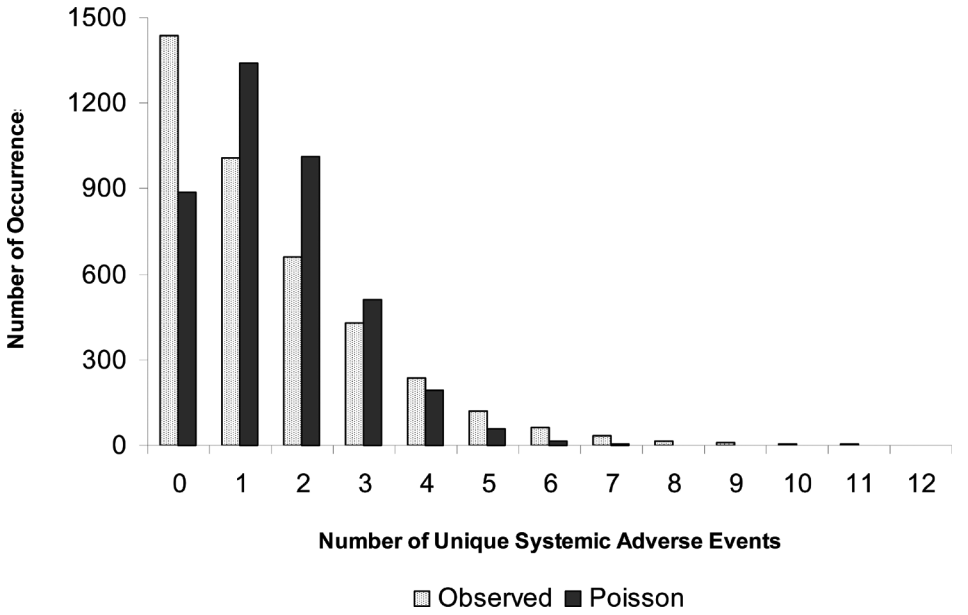

Figure 1 The observed and theoretical Poisson predicted counts for the 1005 study participants (4020 observations) using the systemic adverse event data.

with the theoretical Poisson distribution with a mean of 1.51. It's evident that the independent Poisson model doesn't adequately fit our data but this independent Poisson model ignores the repeated measures aspect of our data. Our primary question is if there's over-dispersion above and beyond a correlated data model, which can't be explained by our potential covariates, and that may be due to zero-inflation and/or unobserved heterogeneity.

Model explanatory variables considered are treatment (Groups I–V), study center, gender, race (black, white, other), and time. Time is defined as the time between a given injection and the first injection. We considered several two way interactions among the covariates. Except for time, all variables were treated as categorical and referent categories for the explanatory variables are Group III, male, white, and Walter Reed Army Institute of Research. In addition, we constructed contrasts to compare Groups I and II, male versus female, and all two-way contrasts for race.

We began by fitting a Poisson model, retaining treatment, study center, gender, and race in the model regardless of significance. The explanatory variable time was significant in the Poisson regression model but all considered two-way interactions were insignificant. Once we obtained our final fitted standard Poisson regression model we fitted the standard NB model using the same covariates. For the zero-inflated/hurdle models we retained all covariates used in the fitted Poisson regression model in both parts of the models (e.g., we retained all covariates in the zero-inflated and Poisson or NB portions of the zero-inflated models).

We computed goodness of fit measures to compare the Poisson, NB, ZIP, ZINB, PH, and NBH models. We compared the Poisson versus NB, ZIP versus ZINB, and PH versus NBH models to determine if the over-dispersion parameter was significant. Formally we tested if the dispersion parameter equals zero, which is on the boundary, using the likelihood ratio test of Self and Liang (1987) and score test for ZIP versus ZINB (Ridout et al., 2001). For all likelihood ratio tests and the score test the dispersion parameter is highly significant (p-value < 0.0001). We tested for excess zeroes by comparing the Poisson and NB models to the ZIP and ZINB models, respectively, using score (Broek, 1995) and likelihood ratio tests. Score and likelihood ratio tests both favored the ZIP and ZINB models (p-values < 0.0001), hence there's evidence of over-dispersion due to excess zeroes. We computed the Vuong (1989) statistic to compare the standard Poisson and ZIP versus PH, and standard NB and ZINB versus NBH models. The Vuong statistic for the PH versus Poisson (10.6, p-value < 0.0001), NBH versus NB (4.3, p-value < 0.0001) overwhelmingly favors the hurdle models. The Vuong statistics for PH versus ZIP (-0.088, p-value $= 0.9299$) and NBH versus ZINB (-0.82, p-value $= 0.4122$) reveals that neither model is favored. Our test results indicate that there is significant over-dispersion due to both heterogeneity and excess zeroes since NB models are preferred over Poisson models and zero-inflated/hurdle models are preferred over the standard Poisson and NB models.

Next we examined the fit of our models using AIC and BIC and compared expected probabilities and resulting counts for each model with observed counts (Table 1). The AIC criterion favors the ZINB and NBH models over all other considered models; however, there is little difference in the ZINB (13162) and NBH (13170) computed AIC values. In contrast, the BIC criterion favors the NB model but the ZINB and NBH models BIC values (BIC = 13332 and 13340, respectively)

Table 1 The actual and predicted frequencies by model and goodness of fit results for fitted models. The actual frequencies correspond to 4020 observed systemic adverse events for four injection for each of the 1005 study participants

Count	Actual*	Poisson	ZIP	PH	NB	ZINB	NBH
0	1437	978.4	1436.4	1436.8	1398.2	1439.4	1437.5
1	1010	1299	810.6	811.2	1089.6	995.5	1000.3
2	660	937.8	789.6	788.9	673.2	688.6	687.7
3	428	489.7	529.5	528.8	385.1	413.3	411.9
4	236	206.9	274.6	274.5	213.7	230.4	229.6
5	122	74.9	117.3	117.5	117.2	123.0	122.8
6	62	24.0	42.9	43.1	64.1	64.0	64.0
7	34	6.9	13.8	13.9	35.2	32.7	32.8
8	14	1.8	4.0	4.01	19.4	16.5	16.6
9	8	0.4	1.0	1.1	10.7	8.3	8.4
10	4	0.1	0.3	0.8	6.0	4.4	4.4
11	4	0.02	0.06	0.06	3.3	2.2	2.2
12	1	0.004	0.01	0.01	1.9	1.1	1.1
Parameters		13	26	26	14	27	27
Pearson's χ^2		1901.5	570.0	569.6	17.9	4.2	4.1
χ^2 p-value		<0.0001	<0.0001	<0.0001	0.1176	0.9804	0.9820
BIC		14033	13549	13550	13303	13332	13340
AIC		13951	13386	13386	13215	13162	13170

Note: Count = number of unique systemic adverse events experienced after an AVA injection. ZIP = zero-inflated Poisson, PH = Poisson hurdle, NB = Negative Binomial, ZINB = zero-inflated Negative Binomial, and NBH = Negative Binomial hurdle models. All models include treatment, gender, race, location, and time covariates.

are close to the NB (13303) model value. The observed number of recorded systemic adverse events that occurred for the 4020 observations, corresponding to the 1005 study participants, is given in Table 1. Predicted counts reveal the standard Poisson model is misspecified. Whereas the NB model predicts the observed frequencies for counts greater than zero adequately but there are some unexplained zeroes. The number of zero counts predicted by the zero-inflated (ZIP and ZINB) and hurdle (PH and NBH) models are close to the observed number of zeroes. However, the ZINB and NBH models fit better than the ZIP and PH models for all other count categories. The computed Pearson's chi-square statistic indicates that the Poisson ($\chi^2 = 1901.5$), ZIP ($\chi^2 = 570.0$), and PH ($\chi^2 = 569.6$) models exhibit lack of fit (p-values < 0.0001). The NB model does not exhibit lack of fit (p-value $= 0.1176$) but the ZINB (p-value $= 0.9804$) and NBH (p-value $= 0.9820$) models substantially improve the fit. The predicted mean for each model is close to the empirical expected value of 1.51 but the variance predicted by the four models varies substantially from the empirical variance estimate of 2.90. Predicted variances are Poisson (1.71), ZIP (2.37), PH (2.38), NB (2.97), ZINB (2.89), and NBH (2.93).

To illustrate the fit of the models by gender we computed the expected frequencies by systemic adverse event count categories (Figs. 2 and 3). Clearly the Poisson model illustrates a lack of fit in that it predicts counts by gender extremely poorly for systemic adverse event count categories 0, 1, and 2. The ZIP and PH models improve the prediction of counts by gender relative to the standard

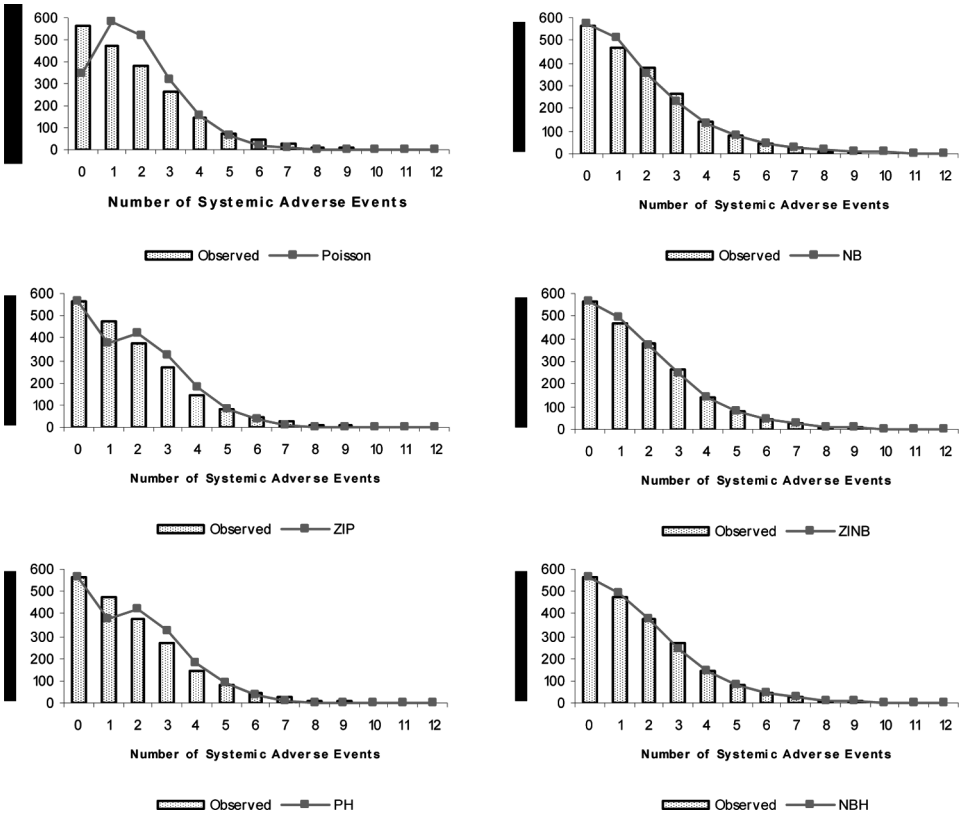

Figure 2 The frequency and predicted frequencies for the number of unique systemic adverse events for the Poisson, Negative Binomial (NB), zero-inflated Poisson (ZIP), zero-inflated NB (ZINB), Poisson hurdle (PH), NB hurdle (NBH) models using the 500 female interim analysis study participants ($N = 2000$ observations).

Poisson but both illustrate a lack of fit, especially for systemic adverse event counts categories 1 and 2. The NB model significantly improves the fit by gender relative to the standard Poisson, ZIP, and PH models, and exhibits no notable lack of fit. However, the ZINB and NBH models both improve the fit by gender relative to the standard NB model. We computed the Pearson's chi-square statistic by grouping all count categories eight and above because of small observed frequencies for these count categories. Pearson's chi-square statistic by gender indicates that the Poisson, ZIP, and PH models exhibit lack of fit for both genders (all p-values < 0.0001). For females, the NB (p-value $= 0.027$) exhibits some lack of fit whereas the ZINB (p-value $= 0.819$) and NBH (p-value $= 0.811$) exhibit no lack of fit. For males, NB (p-value $= 0.075$), ZINB (p-value $= 0.398$), and NBH (p-value $= 0.416$) models reveal no lack of fit but the ZINB and NBH models substantially improve the fit relative to the NB model.

Here the ZINB and NBH models are preferred using all considered criteria except BIC, which slightly favors the NB model. However, given our sample size, the BIC invokes a heavier penalty per parameter, and here for demonstration purposes

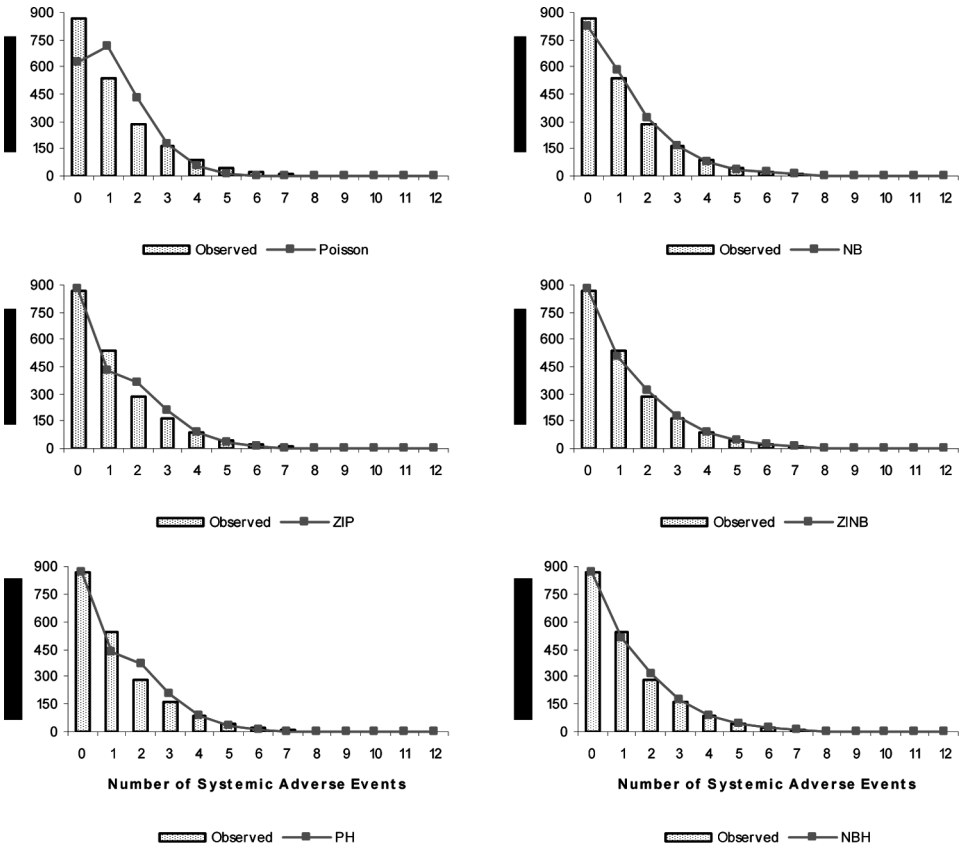

Figure 3 The frequency and predicted frequencies for the number of unique systemic adverse events for the Poisson, Negative Binomial (NB), zero-inflated Poisson (ZIP), zero-inflated NB (ZINB), Poisson hurdle (PH), NB hurdle (NBH) models using the 505 male interim analysis study participants ($N = 2020$ observations).

we did not remove nonsignificant parameters from our models. There is no clear quantitative advantage between the ZINB and NBH models for our data; however, we are modeling the occurrence of systemic adverse events for which all participants are assumed to be at-risk. Hurdle models are generally preferred over zero-inflated models for this type of endpoint since it's assumed we have no structural zeroes. Therefore for our endpoint the NBH is our chosen model. All further results will focus on the NBH model and use other models as comparisons.

Estimated parameters, robust standard errors, and p-values are presented for the NB and NBH models (Table 2). Results for the Poisson, PH, ZIP, and ZINB are not presented for brevity. The NB estimated parameters and robust standard errors are similar to the Poisson model. While some parameters in individual models are not significant, the Wald type test for the effect was significant. For example, a Wald type test for race in the Poisson model revealed it was a significant factor even though the race "other" is not significantly different from the referent race white. We chose not to combine insignificant categories within an explanatory variable.

Table 2 The estimated parameters, robust standard errors (SE), and associated *p*-values for the Negative Binomial and Negative Binomial hurdle (NBH) models using all 4020 observations for the 1005 participants. Referent categories are Group III, male, white, and Walter Reed Army Institute of Research

	Negative Binomial Model			*NBH Model*		
	Estimate	Robust SE	*p*-value	Estimate	Robust SE	*p*-value
Hurdle						
Intercept	NA			0.4112	0.1241	0.0002
Female	NA			0.6345	0.1110	0.0000
Other	NA			0.2013	0.2964	0.2119
Black	NA			−0.3897	0.1444	<0.0001
Group I	NA			0.0764	0.1625	0.4330
Group II	NA			0.0960	0.1144	0.3294
Group IV	NA			−0.7578	0.2535	<0.0001
Group V	NA			−0.5865	0.2519	<0.0001
Baylor	NA			−0.2243	0.1312	0.0467
Emory	NA			0.7197	0.2218	<0.0001
Mayo	NA			0.2312	0.1591	0.0581
UAB	NA			0.1677	0.1906	0.1242
Time	NA			−0.0015	0.0007	0.0008
Negative Binomial						
Intercept	0.2033	0.0599	0.0007	0.3055	0.0867	<0.0001
Female	0.4172	0.0359	<0.0001	0.3228	0.0555	<0.0001
Other	0.1068	0.0753	0.1562	0.0636	0.1367	0.5130
Black	−0.2712	0.0507	<0.0001	−0.2159	0.0740	0.0012
Group I	−0.0465	0.0493	0.3463	−0.1152	0.0751	0.0530
Group II	0.0344	0.0464	0.4589	0.0069	0.0453	0.9055
Group IV	−0.3827	0.0770	<0.0001	-0.1212	0.1897	0.1838
Group V	−0.2809	0.0805	0.0005	−0.0929	0.1528	0.2956
Baylor	−0.0762	0.0651	0.2418	0.0305	0.0703	0.7020
Emory	0.3734	0.0660	<0.0001	0.2588	0.0651	0.0016
Mayo	0.1164	0.0660	0.0780	0.0659	0.0704	0.4124
UAB	0.1229	0.0591	0.0377	0.1182	0.1059	0.1141
Time	−0.0006	0.0002	0.0152	−0.00003	0.0003	0.9049
Dispersion	1.8624	0.1100		2.4925	0.6189	

For the hurdle portion of the NBH model the parameter estimates are in terms of the probability of an event occurring. The hurdle portion of the NBH revealed that the estimates, robust standard errors, and associated *p*-values are similar and there are no discrepancies with the PH model conclusions.

We present contrasts of interest for the NBH, PH, NB, and Poisson models (Table 3). A contrast of interest is Group I versus II and results are similar for the NB and Poisson models for the Group I versus Group II with both having an insignificant RR of approximately 0.92. For other contrasts of interest the NB and Poisson models exhibit similar results with no discrepancies in the conclusions. Both models reveal that females have an elevated risk (RR $\cong$ 1.5 for both models) and that blacks have a reduced number of systemic adverse events relative to whites and other races.

Table 3 Results for selected contrasts using the Negative Binomial hurdle (NBH), Poisson hurdle (PH), Possion, and Negative Binomial models

NBH Model						
Hurdle	**Estimate**	**SE**	***p*-value**	**OR**	**LCL**	**UCL**
Group I vs Group II	−0.0196	0.1482	0.8946	0.9806	0.7334	1.3111
F vs M	0.6345	0.1110	<0.0001	1.8861	1.5171	2.3448
Black vs White	−0.3897	0.1444	0.0070	0.6773	0.5103	0.8988
Black vs Other	−0.5910	0.3091	0.0559	0.5538	0.3021	1.0151
Other vs White	0.2013	0.2964	0.4971	1.2230	0.6839	2.1869
NB				RR	LCL	UCL
Group I vs Group II	−0.1221	0.0693	0.0781	0.8851	0.7726	1.0139
F vs M	0.3228	0.0555	<0.0001	1.3810	1.2386	1.5398
Black vs White	−0.2159	0.0740	0.0035	0.8058	0.6970	0.9316
Black vs Other	−0.2795	0.1464	0.0563	0.7562	0.5675	1.0076
Other vs White	0.0636	0.1367	0.6418	1.0657	0.8151	1.3933
PH model						
Hurdle	Estimate	SE	*p*-value	OR	LCL	UCL
Group I vs Group II	−0.0257	0.1400	0.8543	0.9746	0.7407	1.2825
F vs M	0.6364	0.0795	<0.0001	1.8897	1.6169	2.2085
Black vs White	−0.4016	0.1066	0.0002	0.6692	0.5430	0.8248
Black vs Other	−0.6115	0.2020	0.0025	0.5425	0.3652	0.8061
Other vs White	0.2099	0.1854	0.2576	1.2336	0.8577	1.7742
Poisson				RR	LCL	UCL
Group I vs Group II	−0.1103	0.0679	0.1044	0.8955	0.7839	1.0231
F vs M	0.2860	0.0406	<0.0001	1.3311	1.2292	1.4414
Black vs White	−0.1926	0.0550	0.0005	0.8248	0.7404	0.9188
Black vs Other	−0.2525	0.0895	0.0048	0.7769	0.6518	0.9260
Other vs White	0.0599	0.0781	0.4437	1.0617	0.9109	1.2374
NB Model	Estimate	SE	*p*-value	RR	LCL	UCL
Group I vs Group II	−0.0808	0.0591	0.1717	0.9224	0.8214	1.0357
F vs M	0.4172	0.0359	<0.0001	1.5177	1.4147	1.6282
Black vs White	−0.2712	0.0507	<0.0001	0.7625	0.6904	0.8421
Black vs Other	−0.3780	0.0854	<0.0001	0.6852	0.5796	0.8101
Other vs White	0.1068	0.0753	0.1562	1.1127	0.9600	1.2898
Poisson Model	Estimate	SE	*p*-value	RR	LCL	UCL
Group I vs Group II	−0.0853	0.0582	0.1427	0.9182	0.8193	1.0292
F vs M	0.4025	0.0358	<0.0001	1.4956	1.3941	1.6044
Black vs White	−0.2655	0.0505	<0.0001	0.7668	0.6946	0.8466
Black vs Other	−0.3697	0.0837	<0.0001	0.6909	0.5864	0.8141
Other vs White	0.1042	0.0740	0.1591	1.1098	0.9600	1.2831

Notes: Estimate is the contrast estimate, SE is the robust standard error, OR = odds ratio, RR = risk ratio, LCL and UCL are lower and upper 95% confidence limits.

Contrast results for the NBH model should be evaluated in two parts: 1) hurdle (odds ratio) and 2) conditional NB (risk ratio). The PH and NBH models have similar results for all contrasts. The NBH model reveals that there is virtually no difference in the odds of a systemic adverse event in Group I versus Group II participants (OR = 0.98, p-value = 0.8716). In contrast, once a participant experiences a systemic adverse event there appears to be a lower risk of experiencing more systemic adverse events in Group I (RR = 0.89, p-value = 0.0780) versus Group II. The NBH model hurdle portion exhibits no significant difference in the number of systemic events experienced by whites and other races but blacks are less likely to experience a systemic adverse event relative to the white and other races. In addition, conditional on participants experiencing a systemic adverse event, blacks have a lower risk of experiencing more systemic adverse events than whites and others. The hurdle portion of the NBH model reveals that females are about 1.9 times more likely than males to experience a systemic adverse event after an AVA dose. However, given that females experience a systemic adverse event the conditional NB portion reveals that they have about a 1.4 increased risk versus males of experiencing more systemic adverse events.

6. DISCUSSION

Our study illustrates, for our data, the superiority of standard, zero-inflated, and hurdle negative binomial models over the standard, zero-inflated, and hurdle Poisson models. In particular, we highlight the flexibility of the zero-inflated and hurdle NB models for modeling data characterized by unobserved heterogeneity and/or excess zeroes. Generally, standard Poisson regression is abandoned when over-dispersion is present and NB regression becomes the model of choice. Although the NB model can account for over-dispersion from unobserved heterogeneity and/or positive within period temporal dependency it is unable to account for over-dispersion resulting from excess zeroes. Zero-inflated/hurdle Poisson models can account for over-dispersion resulting only from excess zeroes. Zero-inflated/hurdle NB models are the most flexible of the considered models because they can account for over-dispersion from excess zeroes, unobserved heterogeneity, and/or positive within period temporal dependency.

Our fitted models revealed that our zero-inflated and hurdle models are indistinguishable with respect to goodness of fit measures. However, choosing between the zero-inflated and hurdle models, assuming the Poisson and NB are inadequate because of excess zeroes, should generally be based on study endpoints and goals. If our goal is to develop a prediction model then it's not important which modeling framework we use (i.e., zero-inflated or hurdle), assuming predictions are indistinguishable. In contrast, if our goal is inference then we would suggest it's important to choose the model that is most appropriate given the study design. For example, if we were to ask the question "how many days did you use aspirin in the last 30 days?" to a random group of participants then we would expect to have some participants who never use aspirin (structural zeroes) and some participants who use aspirin but didn't in the last 30 days (sampling zeroes). This naturally leads to zero-inflated modeling framework. In contrast, if we restrict our sample to participants who use aspirin then it naturally leads to the hurdle modeling framework because

we would only have sample zeroes. However, count data containing only sample zeroes may sufficiently be modeled using the standard Poisson or NB models.

Results suggest our data exhibit more over-dispersion from unobserved heterogeneity rather than excess zeroes. This is suggested because our model fit improved more when moving from a Poisson to a NB rather than from a Poisson to a ZIP. However, improvement in model fit from the NB to the zero-inflated/hurdle models suggests there is substantial over-dispersion present from excess zeroes after accounting for unobserved heterogeneity. Our estimated over-dispersion parameter is larger for the NBH (2.49) model than the NB (1.86) model (Table 3). This suggests that over-disperion is greater among the positive counts alone (hurdle model) than when including the zeroes with positive counts in the NB model. Our data exhibits little variability among the proportions of participants experiencing systemic adverse events by dose. This suggests that there is minor heterogeneity due to within period temporal dependency. Generally, in vaccine studies, systemic adverse events are somewhat "rare" and looking at counts of systemic adverse events is likely to lead to excess zeroes and/or unobserved heterogeneity. Our expected counts from the Poisson model revealed a gross underestimation of zero counts and a gross overestimation of the small counts greater than zero, which is consistent with some previous studies (Gray, 2005; Long, 1997). Estimates from the zero-inflated and hurdle models are similar using our data but further investigation is warranted to understand the consequences of using a zero-inflated model rather than a hurdle model, or vice versa, when one model is more natural given the public health endpoint being measured.

Note to reader: At the time of this study Stata was one of the few software packages that could fit zero-inflated models as a standard option, e.g., SAS didn't have a procedure for fitting these models. However, both zero-inflated and hurdle models can be fit in SAS using PROC NLMIXED by specifying the log-likelihood (programs available upon request).

AVRP WORKING GROUP

Wendy A. Keitel, M.D., Baylor College of Medicine

Harry Keyserling, M.D., Emory University

Janiine Babcock, MD, Walter Reed Army Institute of Research, Department of Clinical Trials

Gregory Poland, M.D., Vaccine Research Group and Department of Pediatric and Adolescent Medicine, Mayo Clinic

Robert M. Jacobson, M.D., Chair and Professor of Pediatrics, Mayo Clinic College of Medicine, Mayo Clinic Vaccine Research Group, Mayo Clinic

Scott Parker, M.D., Department of Medicine, University of Alabama at Birmingham, Alabama Vaccine Research Clinic

Nina Marano, DVM, MPH, Dipl ACVPM, Associate Director for Veterinary Public Health, Division of Bacterial and Mycotic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia

Michael McNeil, M.D., Bacterial Vaccine-Preventable Diseases Branch, Epidemiology and Surveillance Division, National Immunization Program, Centers for Disease Control and Prevention, Atlanta, Georgia

Conrad Quinn, PhD, Meningitis and Special Pathogens Branch, Division of Bacterial and Mycotic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia

John Stamper, BA, Office of Informatics, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia

Funding for this project provided by the Department of Health and Human Services-Centers for Disease Control and Prevention.

ACKNOWLEDGMENTS

The authors would like to express our gratitude to the AVA Human Clinical Trial Principal Investigator, Dr. Nina Marano. Without her encouragement and guidance our work would not have come to fruition. We would like to thank Drs. Drew Baughman and Mike McNeil along with two anonymous reviewers for their review of the manuscript.

REFERENCES

Berk, K. N., Lachenbruch, P. A. (2002). Repeated measures with zeros. *Stat. Methods Med. Res.* 11:303–316.

Bohning, D., Dietz, E., Schlattmann, P., Mendonca, L., 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(2):195–209.

Broek, J. V. D. (1995). A score test for zero-inflation in a Poisson distribution. *Biometrics* 51:738–743.

Brown, C. J., Pagan, J. A., Rodriguez-Oreggia, E. (2005). The decision-making process of health care utilization in Mexico. *Health Policy* 72(1):81–91.

Cameron, A. C., Trivedi, P. K. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

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

Chin, H. C., Quddus, M. A. (2003). Modeling count data with excess zeroes – an empirical application to traffic accidents. *Sociol. Meth. Res.* 32(1):90–116.

Dobbie, M. J., Welsh, A. H. (2001). Modelling correlated zero-inflated count data. *Australian and New Zealand Journal of Statistics* 43:431–44.

Gray, B. R. (2005). Selecting a distributional assumption for modelling relative densities of benthic macroinvertebrates. *Ecological Modelling* 185:1–12.

Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics* 56:1030–39.

Hall, D. B., Zhang, Z. (2004). Marginal models for zero-inflated clustered data. *Statistical Modelling* 4:161–180.

Krobot, K. J., Kaufman, J. S., Christensen, D. B., Preisser, J. S., Miller, W. C., Ibrahim, M. A. (2005). Accessing a new medication in German: a novel approach to assess a health insurance-related barrier. *Ann Epidemiol.* 15(10):756–761.

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

Liu, T.-C., Chen, C.-S. (2002). An analysis of private health insurance purchasing decisions with national health insurance in Taiwan. *Social Science and Medicine* 55:755–774.

Long, J. S. (1997). Count outcomes: regression models for counts. In: Long, J. S., ed. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications, pp. 217–250.

Min, Y., Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling* 5:1–19.

Moulton, L. H., Curriero, F. C., Barroso, P. F. (2002). Mixture models for quantitative HIV RNA data. *Stat. Methods Med. Res.* 11:317–325.

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

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

Self, S. G., Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *J. Am. Stat. Ass.* 82:605–610.

Sheu, M. L., Hu, T. W., Keeler, T. E., Ong, M., Sung, H. Y. (2004). The effect of a major cigarette price change on smoking behavior in California: a zero-inflated negative binomial model. *Health Econ.* 13(8):781–91.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57(2):307–333.

Yau, K. K. W., Lee, A. H. (2001). Zero-inflated Poisson regression with random effects to evaluate an occupations injury prevention programme. *Stat. Med.* 20:2907–2920.