Purpose:

The purpose of this paper is two-fold. First, to develop a model that best explains variability in number of school days suspended. Using Occam's razor in model comparison and selection, this paper examined Poisson (assumes no zero-inflation or overdispersion), negative binomial (NB) (assumes no zero-inflation), Poisson hurdle (PH) (assumes no overdispersion), and negative binomial hurdle (NBH) models to determine which model provides the best fit to the data, has the highest predictive validity, and cross-validates the best. Second, to examine the consequences, if any, of ignoring zero-inflation and overdispersion in the context of the school suspension data. In other words, if a certain probability model provides the best fit, what are the consequences of selecting an alternative model? Contingent on the results from the first aim, this paper will explore to what extent prediction, statistical significance, and overall inferences change if zero-inflation and overdispersion are ignored.

Theoretical Background:

Count data frequently arise in educational settings. For example, the number of days a student is suspended in a school year is a discrete, non-negative count variable. Generally, count data are considered realizations of the Poisson model (Agresti, 2002). However, there are two major problems when conceptualizing school days suspended as a realization of the Poisson model: zero-inflation and overdispersion. These problems arise because the Poisson model is often too simplistic of a model in an applied setting. In the Poisson model, the E(Y) = Var(Y) = λ and this equality is easily and frequently violated.

The first problem, zero-inflation, occurs when there is an excess of zeros in the observed data and more zeros are observed than would be expected if the data-generating process was a Poisson. When studying school suspensions, the majority of students are never suspended and this results in a large point mass at zero (Figure 1). Subsequently, the mean will be pulled towards the zero count. If zero-inflation is not corrected for, then there will be unreasonable fit for both the zeros and non-zero counts and this may lead to parameter estimates and standard errors that are biased (Atkins & Gallop, 2007; Angers & Biswas, 2003; Bohning, 1998; Lambert, 1992; Zuur, Ieno, Walker, Saveliev, & Smith, 2009).

The second problem, overdispersion, occurs when there is greater variability in the data than would be expected from a Poisson model. Dispersion is defined as Var(Y)/E(Y) (Mullahy, 1986). In the Poisson model, the Var(Y) = E(Y) = λ, therefore overdispersion is said to occur when this ratio is larger than 1. If overdispersion is ignored then the standard errors for the parameter estimates will be seriously underestimated, leading to overly optimistic standard errors and p-values, and lack-of-fit based on deviance tests (Hinde & Demetrio, 1998; Agresti, 2002; Atkins & Gallop, 2007; Hall, 2000; Lambert, 1992; Yau, Wang, & Lee, 2003).

In the context of school suspensions, the majority of students will not be suspended. However, there will also be a sizable minority of students that are suspended numerous times. Using a Poisson model to explain this data may result in inadequate fit and incorrect inferences.

Methods:

The data were randomly split into a training and a test set. The training set consisted of approximately 66% of the data and was used to build the models and obtain parameter estimates. The test set was then used to determine how well the model fit the withheld data. Using the training data, I examined and fit four models: Poisson, NB, PH, and NBH models. The models all contained main effects only and there were no interactions between any of the factors in these models.

To compare models and examine relative model fit, I used Vuong's test (Vuong, 1989) and Akaike's information criterion (AIC) (Burnham & Anderson, 2004). Under the null hypothesis, Vuong's statistic follows a standard normal and one chooses a critical value, c, for a significance level (often 1.96 to correspond to $\alpha = 0.05$). In this paper c was selected as 1.96 to be consistent with standard applied statistical convention. Models with the lowest AIC are the best fitting and favored models.

To examine absolute model fit, I plotted the predicted curve against the observed data and examined how similar the two curves were. I then compared the predicted zeros from each model against the observed zero count. This was done for both the training and test data sets.

All analyses were performed in R (version 2.15.0) (R Development Core Team, 2012).

Data Source:

Data were collected from the 2003-2004 through 2007-2008 school years on 13,606 students in grade 8 in a large, urban school district. The dependent variable in this study was total days suspended.

Students were divided into one of five risk groups: Homeless or highly mobile (HHM) more than one year during the study (HHM>1), HHM only one year during the study (HHM=1), receiving free-priced meals but not HHM (FREE), receiving reduced-priced meals but not HHM (RED), or the non-low income general group (GEN). The HHM students were broken into two separate groups as it was believed that multiple years of homelessness and high mobility represented a greater risk (a chronic risk) than just an episode of homelessness (a more acute risk). Additional data on gender, ethnicity, English language learner status (ELL), and special education status were provided by the school district.

Results:

Table 1 presents the results from the model comparison using the AIC and Vuong's statistic on the training data. Table 1 shows overwhelming support for the NBH model best on the AIC and Vuong's statistic. We see that the NB fit better than the PH model and the PH model fit better than the Poisson. Vuong's test found that the NBH was favored over the PH, NB, and Poisson model ($V > 14$, $p < .001$ for all comparisons); that the NB was favored over the PH and the Poisson model ($V > 10$, $p < .001$ for both comparisons); and finally the PH was favored over the Poisson model ($V = 33.58$, $p < .001$). This suggests that overdispersion may be the bigger problem with this dataset as the model that accounted solely for overdispersion (the

NB model) fit better than the model that accounted solely for zero-inflation (the PH model). However, it is clear that there is strong evidence of both overdispersion and zero-inflation conditional on the factors examined.

Table 2 shows the parameter estimates for the NBH model. For the count component, i.e. parameter estimates conditional on a student being suspended, we can see that the mean number of school days suspended for African Americans is 1.58 times greater than for non-Hispanic White students and that the mean number of school days suspended for American Indians is 1.46 times greater than for non-Hispanic White students, holding the other variables constant. We see there is no difference between non-Hispanic White students and Asian American or Hispanic students in the mean number of school days suspended. The expected number of school days suspended for RED students is .66 times greater than that for HHM>1 students and for GEN students is .53 times greater than that for HHM>1 students, holding the other variables constant. There is no difference in the mean number of school days suspended for HHM> 1 and HHM=1 or FREE. The expected number of school days suspended for students in special education is 1.34 times greater than for students not receiving special education, holding the other variables constant. The expected number of school days suspended for male students is 1.13 times greater than for female students. There is no difference in the expected mean number of school days suspended based on ELL status.

For the zero component, i.e. parameter estimates for the probability of being suspended, we can see that the expected odds of an African American student being suspended are 3.10 times the odds of a non-Hispanic White student; the expected odds of an Asian American student being suspended are .53 times the odds of a non-Hispanic White student; and the expected odds of an American Indian student being suspended are 2.48 times the odds of a non-Hispanic White student, holding the other variables constant. We see there are no differences in the odds of being suspended between non-Hispanic White and Hispanic students; HHM>1 and HHM=1; and HHM>1 and FREE students. The expected odds of being suspended for a RED student are .62 times the odds for a HHM>1 student and the expected odds of being suspended for a GEN student are .26 times the odds for a HHM>1 student, holding the other variables constant. The expected odds of being suspended for a student receiving special education services are 1.46 times the odds of a student not receiving special education services, holding the other variables constant. There are no differences in the odds of being suspended conditional on being an ELL student. Finally, the expected odds of being suspended for a male student are 1.88 times the odds of a female student, holding the other variables constant.

Comparing the NBH to the Poisson models, we find that the Poisson models have smaller standard errors and p-values (not presented) than the NBH model. Selecting a Poisson model then would result in erroneous conclusions about the significance of variables.

Figures 2 & 3 shows the predicted number of school days suspended for the four probability models examined using the training data and testing data, respectively. We can see that only the PH and NBH models (the purple x's and magenta triangles) properly modeled the zero count and that the Poisson and NB models (green circles and oranges crosses) predicted substantially fewer zeros than those observed. We see that the Poisson and NB models perform poorly. We can also see that the PH fails to predict subjects with greater than 17 days of school suspended. In contrast, the NBH model performs reasonably well for all counts with the exception of 3 days suspended where the NBH under predicts the observed count.

Significance:

Studying number of days that students are suspended from school is an inherently intractable statistical problem. Many students are never suspended while a sizable minority of students are suspended for more than a week. These non-offenders and chronic offenders increase the number of observed zeros and cause the variance to become large relative to the mean. Traditional count models, such as the Poisson or NB model, might fail to accommodate for the zero-inflation and/or overdispersion that is inherently present in data of this nature.

There are several practical implications of this study. First, considering the Poisson model to model discrete counts over a time interval (an academic school year) seems reasonable, but failing to properly accommodate zero-inflation and overdispersion would result in wrong conclusions about parameter estimates. In the Poisson model, all but the comparison between HHM>1 and HHM=1 were significant. Furthermore, we can see that after modeling out the zeros (which occurs with the zero component in the hurdle models) and focusing solely on those students that are suspended, the data are still overdispersed conditional on the examined covariates. In fact, the conclusions about the parameters from the count component for the NBH model agree largely with the NB model (not presented).

Ignoring the overdispersion when examining school suspensions would result in concluding significant differences in school days suspended associated with the aforementioned factors. Given that an educator in a school would be interested in factors associated with how many school days a student might be expected to be suspended using a Poisson instead of a NBH would result in practitioners focusing on unimportant factors and wasting time, money, and resources if these factors were used to develop an intervention.

Second, researchers and educators studying school suspensions might be interested in knowing not only what factors are associated with the number of school days a student might be expected to be suspended but also what factors are associated with whether a given student would be expected to be suspended. The usefulness of hurdle models is that they allow you to simultaneously answer both of these questions.

References:

Agresti, A. (2002). Categorical data analysis. John Wiley and Sons.

Atkins, D., & Gallop, R. (2007). Rethinking how family researchers model in-frequent outcomes: a tutorial on count regression and zero-inflated models. Journal of Family Psychology, 21(4), 726–735. doi:10.1037/0893-3200.21.4.726.

Angers, J., & Biswas, A. (2003). A Bayesian analysis of zero-inflated generalized Poisson model. Computational Statistics & Data Analysis, 42(1-2), 37–46.

Bohning, D. (1998). Zero-inflated Poisson models and C.A.MAN: A tutorial col-lection of evidence. Biometrical Journal, 40(7), 833–843.

Burnham, K., & Anderson, D. (2004). Multimodel inference: understanding AIC and BIC in model selection. Sociological Methods & Research, 33(2), 261– 304. doi:10.1177/0049124104268644.

Hall, D. (2000). Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics, 1030–1039.

Hinde, J., & Demetrio, C. (1998). Overdispersion: models and estimation. Com-putational Statistics & Data Analysis, 27(2), 151–170.

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

Mullahy, J. (1986). Specification and testing of some modified count data models. Journal of Econometrics, 33(3), 341–365.

R Development Core Team. (2012). R: a language and environment for statistical computing. ISBN 3-900051-07-0. R Foundation for Statistical Computing.

Vienna, Austria. Retrieved from R Foundation for Statistical Computing: http://www.R-project.org/

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica: Journal of the Econometric Society, 307–333.

Yau, K., Wang, K., & Lee, A. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. Biometrical Journal, 45(4), 437–452. doi:0323-3847/03/0406-0437.

Zuur, A., Ieno, E., Walker, N., Saveliev, A., & Smith, G. (2009). Zero-truncated and zero-inflated models for count data. Mixed effects models and extensions in ecology with R, 261–293.

Table 1. Model comparison for the training data based on AIC and Vuong's test.

Model	AIC	Ranking based on Vuong's test
Poisson	43157	4
Negative Binomial	22123	2
Poisson Hurdle	24818	3
Negative Binomial Hurdle	21238	1

Table 2. Summary of the negative binomial hurdle model.

		Estimate	SE	Wald	Pr(>|z|)
Count	Intercept	1.26	.12	10.08	<.001
	AA	.46	.08	5.47	<.001
	Asian	.14	.13	1.05	.29
	Hispanic	.13	.11	1.19	.24
	AI	.38	.11	3.41	<.001
	HHM=1	-.04	.10	-.35	.73
	FREE	-.18	.09	-1.94	.05
	RED	-.42	.14	-3.02	<.001
	GEN	-.63	.12	-5.21	<.001
	Spec Ed	.29	.05	6.36	<.001
	ELL	-.09	.05	-1.85	.06
	Male	.12	.04	2.75	.01
	Log(θ)	.31	.06	5.20	<.001
		Estimate	SE	Wald	Pr(>|z|)
Zero	Intercept	-1.73	.15	-11.28	<.001
	AA	1.13	.09	12.58	<.001
	Asian	-.63	.14	-4.53	<.001
	Hispanic	.15	.12	1.25	.21
	AI	.91	.13	6.96	<.001
	HHM=1	-.00	.14	-.02	.99
	FREE	-.20	.13	-1.57	.12
	RED	-.48	.18	-2.70	.01
	GEN	-1.35	.15	-8.94	<.001
	Spec Ed	.38	.06	6.24	<.001
	ELL	-.05	.06	-.78	.43
	Male	.63	.05	11.89	<.001

Note: AA refers to African American, AI to American Indian, Spec Ed to special education. Count refers to the parameter estimates for the count component and zero refers to the parameter estimates for the zero component, respectively.

Figure 1. Marginal distribution of school days suspended.

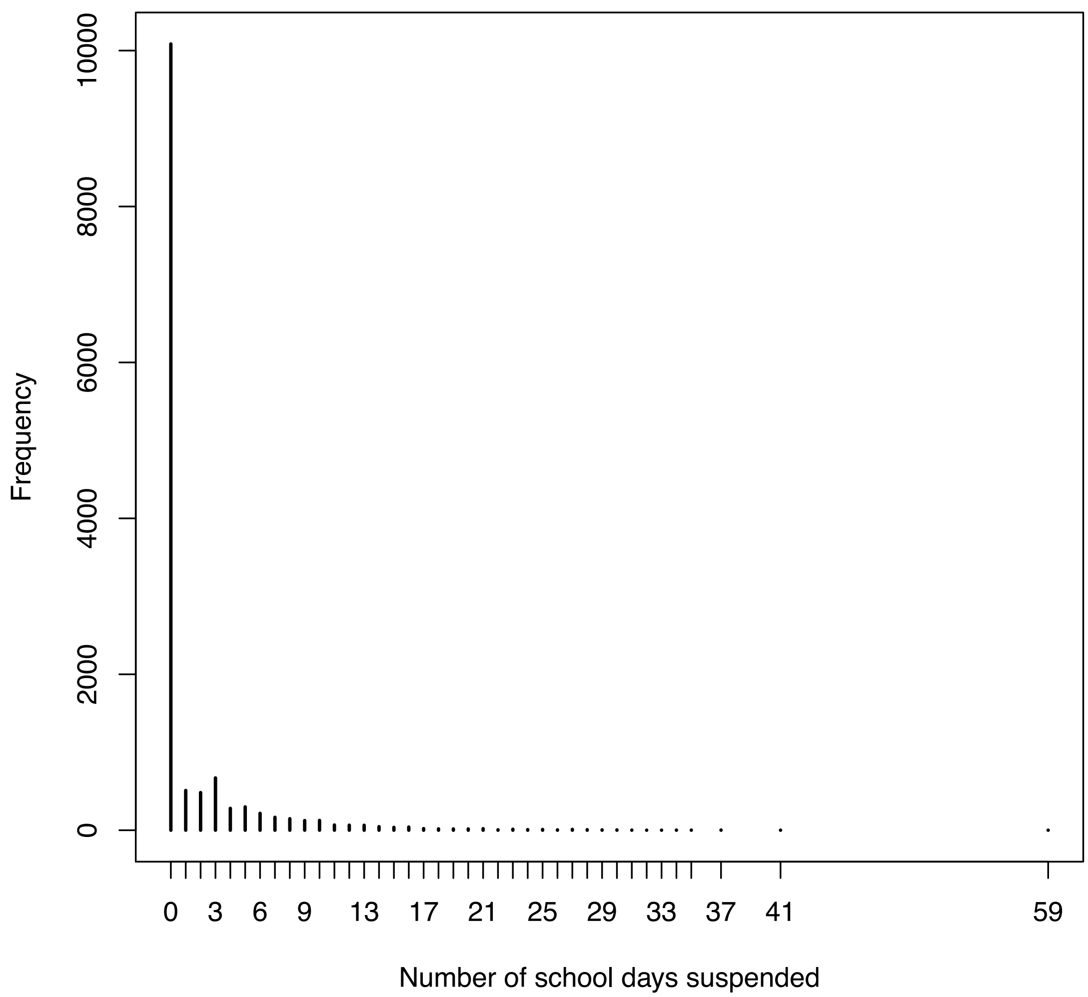

Figure 2. Predicted number of school days suspended using the training data set. The black bars correspond to the observed number of school days suspended, the green circles to the predicted values from the Poisson; the orange crosses to the predicted from the NB; the purple x's to the predicted values from the PH; and the magenta triangles to the predicted values from the NBH.

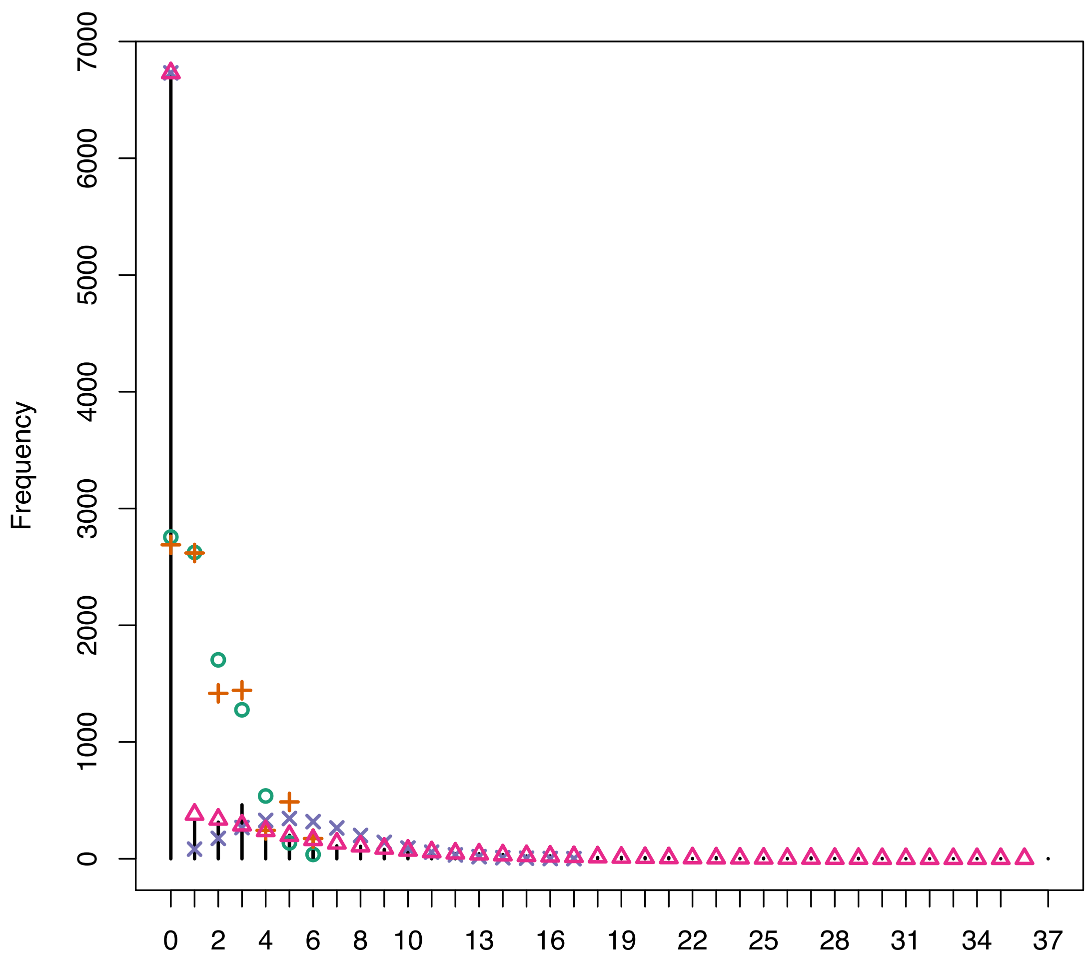

Figure 3. Predicted number of school days suspended using the testing data set. The black bars correspond to the observed number of school days suspended, the green circles to the predicted values from the Poisson; the orange crosses to the predicted from the NB; the purple x's to the predicted values from the PH; and the magenta triangles to the predicted values from the NBH.