Modeling approaches for estimating a treatment effect from count data in the presence of many zeros and over-dispersion

Johnson E., Arbet-Engels C., Nguyen, H., Zeger S.L.

In a data from clinical studies comparing two insulin therapies for the treatment of diabetes, we seek to 1) compare the mean rate of hypoglycemic episodes between the two therapies, after adjusting for glycemic control and 2) estimate the levels of glycemic control for each therapy for a fixed mean rate of hypoglycemic event. The analysis is complicated by a large frequency of zero observations and several large positive counts of episodes. Therefore, standard Poisson regression methods, which assume equality of the mean and variance of the distribution of episodes, are inappropriate.

We implement several modeling approaches to estimate the treatment effect including an over-dispersed Poisson model, zero-inflated Poisson model and two-part model. In the two-part model, we first estimate the probability of a zero observation using a logistic regression and then model the expected mean number of non-zero episodes using an over-dispersed Poisson model. We compare the power to detect a treatment effect between the modeling approaches via simulation.

We invert the regression models to estimate the levels of glycemic control for a fixed mean rate of hypoglycemic episodes. Confidence intervals are obtained via a bootstrap.

Keywords: Over-dispersed poisson, zero-inflated poisson, two-part model, inverse regression