

Modelling Overdispersion and Markovian Features in Count Data

Iñaki F. Trocóniz (1), Elodie Plan (2), Raymond Miller (3), Mats O. Karlsson (2)

(1) Department of Pharmacy and Pharmaceutical Technology, School of Pharmacy; University of Navarra; Pamplona; Spain; (2) Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden; (3) Pfizer Global Research and Development, Ann Arbor, MI 48108, USA

Background: The number of counts (events) per unit of time is a discrete response variable that is generally analyzed with the Poisson distribution (PS) model.^[1] The PS model makes two assumptions: the mean number of counts (λ) is equal to the variance (σ^2) of the data, and the number of counts occurring in non-overlapping intervals of time are independent. However, many counting outcomes show greater variability than predicted by the PS model. This phenomenon is called overdispersion.^[2] Moreover the pharmacometric community is currently realizing that an increasing number of pharmacodynamic variables show interdependency between neighbouring measurements, a feature that has been modelled incorporating Markovian elements.^[3]

Objectives: To implement and explore, in the population context, different distribution models accounting for the overdispersion and the Markovian patterns in the analysis of count data.

Methods: Daily seizures count data obtained from 551 subjects during the 12 weeks screening phase of a double-blind, placebo-controlled, parallel-group multicenter study performed in epileptic patients with medically refractory partial seizures, were used in the current investigation.

The following distribution models were fitted to the data: (1) the PS, (2) the Zero Inflated Poisson (ZIP), (3) the Inverse Binomial (INB), and (4) the Zero Inflated Inverse Binomial (ZINB), models. The Markovian patterns were introduced by estimating different λ s and overdispersion parameters depending on whether the previous day was a seizure or a non-seizure day. All analyses were performed using NONMEM VI.

The following diagnostics were used to explore the model performances: (i) visual predictive checks, where the 25th, 50th, and 75th percentiles of the σ^2 vs λ profiles obtained from more than 10000 simulated patients were computed, and (ii) numerical predictive checks, where the mean (10th-90th percentiles) from 100 simulated datasets were calculated for the following data descriptors: mean number of transitions, mean elapsed time, mean seizure daily counts, and number of daily counts equal to 0, 1, 2, 3, 4, 5, >5 - ≤ 10 , >10 - ≤ 15 and >15.

Results: All models were successfully implemented in NONMEM and all overdispersed models improved the fit with respect to the PS model. The INB model resulted in the most relevant model fit to the data providing a minimum value of the objective function 7275 points lower than the PS model for six extra parameters. Including Markovian patterns in λ and in the overdispersion parameter improved the fit significantly ($P < 0.0001$). The typical population estimates of λ if the previous day was a seizure or a non-seizure day were 0.53 and 0.32, respectively, and values of the overdispersion parameter were 0.15, and 0.58, respectively.

The diagnostics used to evaluate model performance covered the main features of the data and resembled the changes found in the minimum values of the objective function.

Conclusions: The ZIP, INB, and ZINB models were all capable of dealing with the overdispersion in count data and allowed the flexible incorporation of Markovian elements. They provide, in addition to the mean number of counts, additional possibilities to test placebo/drug/disease progression effects such as the degree of overdispersion, and transition probabilities.

References:

- [1] Miller R, Frame B, Corrigan BW, Burger P, Bockbrader H, Garofolo H, Lalonde R. Exposure-response analysis of pregabalin add-on treatment of patients with refractory partial seizures. *Clinical Pharmacology & Therapeutics* **73**: 491-505 (2003)
- [2] Slymen DJ, Ayala GX, Arredondo EM, Elder JP. A demonstration of modelling count data with an application to physical activity. *Epidemiologic Perspectives & Innovations* **3**: 1-9 (2006)
- [3] Karlsson MO, Schoemaker RC, Kemp B, Cohen AF, van Gerven JMA, Tuk B, Peck CC, Danhof M. A pharmacodynamic Markov mixed-effects model for the effect of temazepam on sleep. *Clinical Pharmacology & Therapeutics* **68**: 175-88 (2000)