A Weibull-count approach for handling under- and overdispersed longitudinal/clustered data structures

Martial Luyts **Geert Molenberghs** **Geert Verbeke**
Koen Matthijs **Eduardo Ribeiro Jr** **Clarice Demétrio** **John Hinde**

Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat)

University of Leuven, Belgium

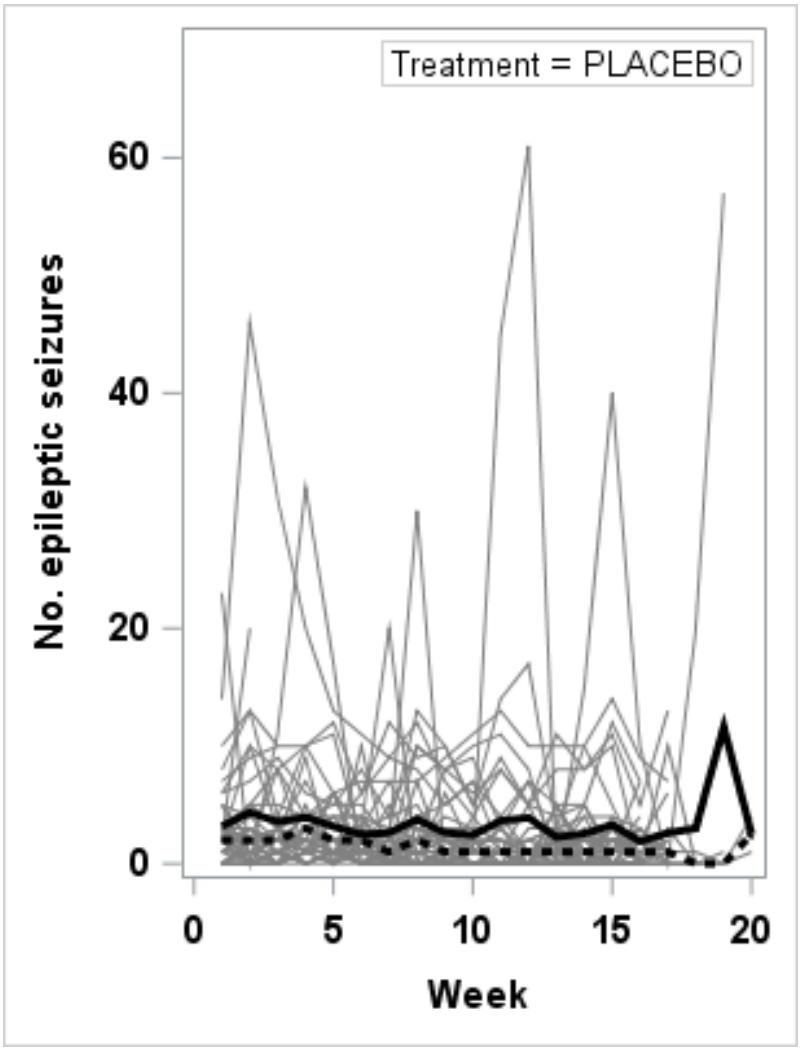
martial.luyts@kuleuven.be

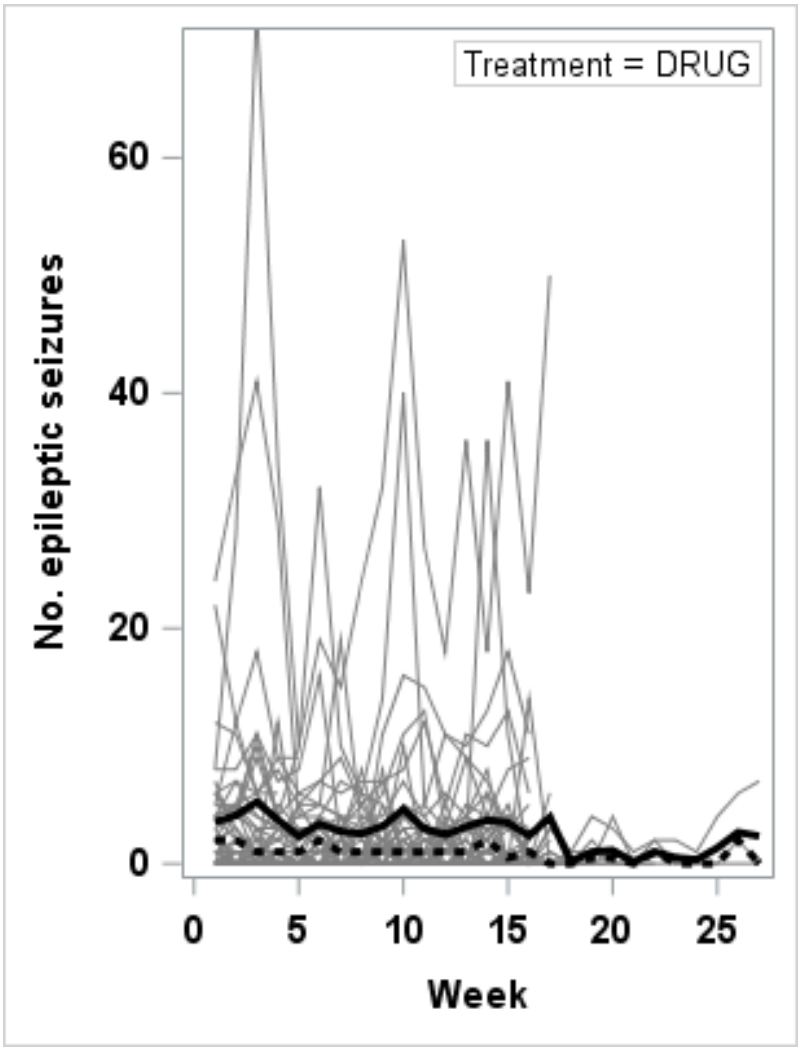
www.ibiostat.be

ENAR, March 26, 2018

1. Clinical trial in epileptic patients

- A **randomized, double-blind, parallel group multi-center study**
- **Aim:** Compare placebo with a new Anti-Epileptic Drug (AED) and one or two other AEDs
- Randomization after 12-weeks stabilization period (45 patients to placebo, 44 to the new treatment)
- Number of epileptic seizures, i.e., a count outcome, measured for up to 27 weeks of follow-up
- **Research interest** is whether or not the additional new treatment reduces the number of seizures


Treatment = PLACEBO
No. epileptic seizures
60
40
20
0
0
5
10
15
20
Week


Treatment = DRUG
No. epileptic seizures
60
40
20
0
0
5
10
15
20
25
Week

2. The Weibull-count approach

- First introduced by Nakagawa and Osaki (1975)
- Recently examined for dispersion by Klakattawi et al. (2018)
- The Weibull-count (DW) approach:
  - Let Y_i, $i = 1, \ldots, n$, be DW distributed
  - **Probability mass function:**

$$P(Y_i = y_i) = q^{y_i^\rho} - q^{(y_i+1)^\rho}$$

$$0 < q < 1, \qquad \rho > 0$$

Special cases:

- $\rho = 1$ & $q = 1 - p \Rightarrow$ Geometric distribution;
- $\rho = 2$ & $q = \theta \Rightarrow$ Discrete Rayleigh (DR) distribution (Roy, 2004);
- $\rho \to +\infty \Rightarrow$ DW approaches a Bernoulli distribution with probability q.

Mean & variance expression:

$$\mathsf{E}(Y_i) \;=\; \mu_{ij} = \sum_{n=1}^{+\infty} \mathrm{e}^{-\lambda_{ij} \cdot n^{\rho}},$$

$$\mathsf{Var}(Y_i) \;=\; 2 \cdot \sum_{n=1}^{+\infty} n \cdot \mathrm{e}^{-\lambda_{ij} \cdot n^{\rho}} - \mu_{ij} - \mu_{ij}^2.$$

Nice property:

Its π-th $(0 < \pi < 1)$ quantile has a closed-form expression, given by

$$\mathsf{Q}_\pi \;=\; \left\lceil \left(\frac{\log(1-\pi)}{\log(q)} \right)^{1/\rho} - 1 \right\rceil .$$

- **Median expression:**

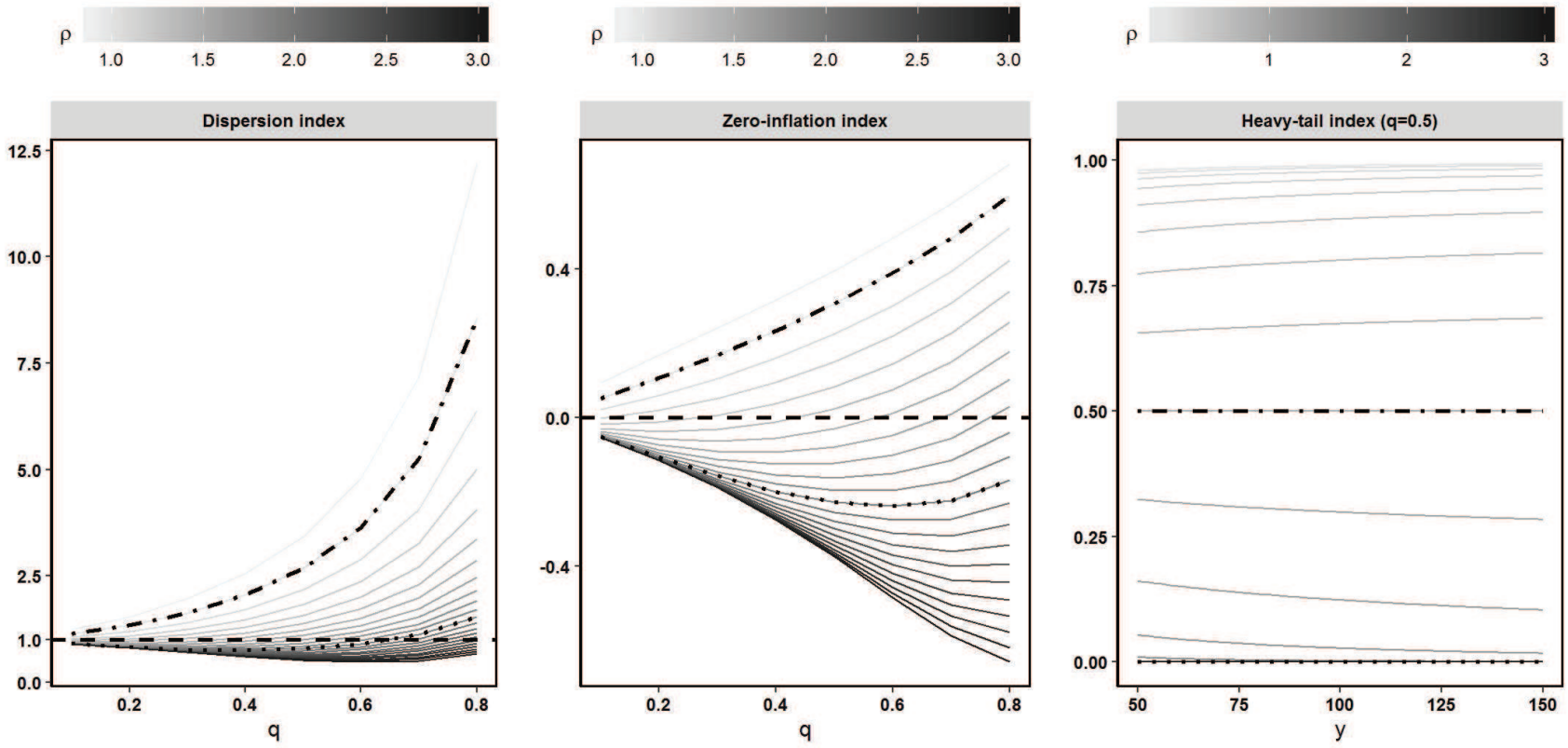
$$\log(\mathsf{Q}_{1/2} + 1) \;=\; \frac{1}{\rho} \cdot \{\log[\log(2)] - \log[-\log(q)]\} .$$

$\Rightarrow$ Regression context:

$$\log[-\log(q_i)] = \mathbf{x}_i^{'} \cdot \boldsymbol{\beta}.$$

2.1. Characteristics

$$\mathsf{DI} = \frac{\mathsf{Var}(Y_i)}{\mathsf{E}(Y_i)}, \qquad \mathsf{ZI} = 1 + \frac{\log P(Y_i=0)}{\mathsf{E}(Y_i)}, \qquad \mathsf{HT} = \frac{P(Y_i=y_i+1)}{P(Y_i=y_i)}, \text{ for } \; y_i \to \infty.$$



(a) DI **(b)** ZI **(c)** HT for q = 0.5

2.2. Adjust for longitudinal/hierarchical structures

- [Epilepsy dataset] Number of epilepsy attacks are recorded over time $\Longrightarrow$ **Longitudinal structure** is present!
- Notation of Y_i now extends to Y_{ij}, which presents the j-th outcome $(j = 1, \ldots, n_i)$ in cluster/subject i $(i = 1, \ldots, N)$
- A mixed effects approach:

$$\log[-\log(q_{ij})] = \boldsymbol{x}_{ij}^{'} \cdot \boldsymbol{\beta} + \boldsymbol{z}_{ij}^{'} \cdot \boldsymbol{b_i},$$

$$\boldsymbol{b_i} \sim N(\mathbf{0}, D)$$

3. Analyzing the Epilepsy dataset

- Let Y_{ij} be the number of epileptic seizures that patient i experiences during week j of the follow-up period

- Let t_{ij} be the time-point at which outcome Y_{ij} has been measured, i.e., $t_{ij} = 1, 2, \ldots$ until at most 27

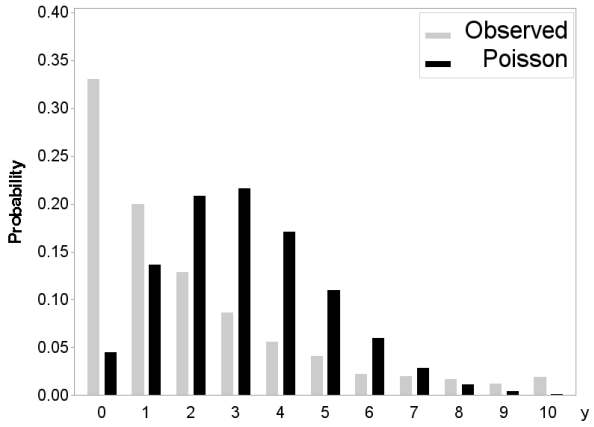
- The hierarchical DW model is considered here, with

$$\text{log}[-\text{log}(q_{ij})] = \begin{cases} (\beta_0 + b_i) + \beta_1 \cdot t_{ij}, & \text{if placebo} \\ (\beta_2 + b_i) + \beta_3 \cdot t_{ij}, & \text{if treated} \end{cases}$$

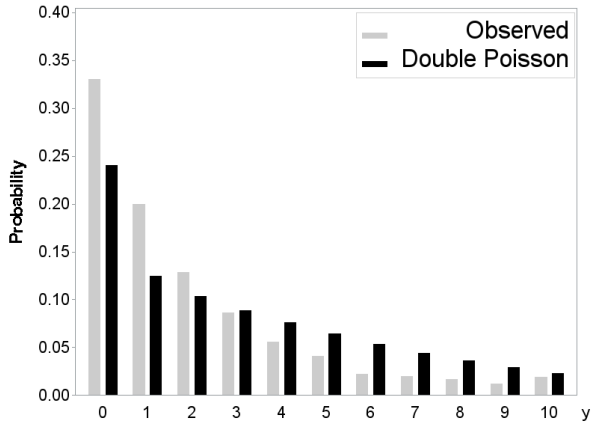
$$b_i \sim N(0, \sigma^2)$$

- Results of the **univariate cases**:

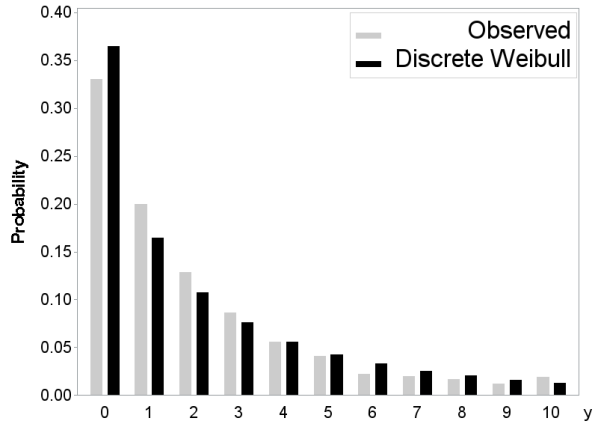
		P	**DP**	**DW**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intercept placebo	β_0	1.2662 (0.0424)	1.2662 (0.1054)	0.7341 (0.1002)
Slope placebo	β_1	−0.0134 (0.0043)	−0.0134 (0.0108)	−0.0174 (0.0095)
Intercept treatment	β_2	1.4531 (0.0383)	1.4531 (0.0953)	0.8278 (0.0992)
Slope treatment	β_3	−0.0328 (0.0038)	−0.0328 (0.0095)	−0.0317 (0.0085)
Difference in slopes	$\beta_3 - \beta_1$	−0.0195 (0.0058)	−0.0195 (0.0144)	−0.0143 (0.0127)
Ratio of slopes	β_3/β_1	2.4576 (0.8480)	2.4576 (2.1094)	1.8189 (1.1027)
	ϕ	−−	0.1616 (0.0061)	−−
	ρ	−−	−−	0.7383 (0.0172)
-2 loglik		11590.0	6815.6	6291.3
AIC		11598.0	6825.6	6301.3
BIC		11619.0	6851.9	6327.6



(a) Poisson model



(b) Double Poisson model



(c) Discrete Weibull model

- Results of the **hierarchical cases**:

		PN	**DPN**	**DWN**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intercept placebo	β_0	0.8179 (0.1677)	0.8314 (0.1721)	1.4322 (0.2182)
Slope placebo	β_1	−0.0143 (0.0044)	−0.0146 (0.0067)	−0.0297 (0.0098)
Intercept treatment	β_2	0.6475 (0.1701)	0.6730 (0.1753)	1.1352 (0.2194)
Slope treatment	β_3	−0.0120 (0.0043)	−0.0129 (0.0065)	−0.0117 (0.0093)
Difference in slopes	$\beta_3 - \beta_1$	0.0023 (0.0062)	0.0018 (0.0093)	0.0180 (0.0135)
Ratio of slopes	β_3/β_1	0.8398 (0.3979)	0.8777 (0.5979)	0.3947 (0.3382)
Std. dev. random effect	σ	1.0755 (0.0857)	1.0458 (0.0875)	1.2646 (0.1062)
	ϕ	−−	0.4355 (0.0169)	−−
	ρ	−−	−−	1.3075 (0.0340)
-2 loglik		6271.9	5652.2	5451.8
AIC		6281.9	5664.2	5463.8
BIC		6294.3	5679.1	5478.7

4. Conclusions

- Able to flexibly model highly overdispersed, zero-inflated, heavy-tailed and correlated data
- Capable of modeling some low overdispersed regions with zero-deflation (e.g., the DR approach) and even underdispersed data
- Interpretations of the parameters can directly be related to the median profile


danke
謝謝
ngiyabonga
teşekkür ederim
спасибо
dank je
thank you
gracias
tapadh leat
mochchakkeram
bedankt
hvala
mauruuru
dziękuję
go raibh maith agat
obrigado
sagolun
sukriya
kop khun krap
arigatō
takk
dakujem
grazie
мерси
terima kasih
감사합니다
ευχαριστώ
merci