Beta-binomial mixed models: Application in longitudinal health related quality of life data

Josu Najera-Zuloaga[1], Dae-Jin Lee[1], Inmaculada Arostegui[1,2]

[1] *Basque Center for Applied Mathematics*
Applied Statistics Research Group

[2] *UPV/EHU - Dept of Applied Maths, Statistics and OR*

basque center for applied mathematics

EXCELENCIA SEVERO OCHOA

September 8, 2016

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

Table of contents

Motivation

Health related quality of life:

- Increasingly important indicator of health status.
- Information about the disease and its impact in the patient in a standardized, comparable and objective way.
- Short-Form 36 Health Survey.

Motivation

Health related quality of life:

- Increasingly important indicator of health status.
- Information about the disease and its impact in the patient in a standardized, comparable and objective way.
- Short-Form 36 Health Survey.

Goal: Study the relationship of HRQoL with patients and diseases characteristics.

Short-Form 36 Health Survey

- It has 36 items, with different answer options.
- It provides a way to measure HRQoL from the patient's point of view.
- Represents eight of the most important health dimensions:

1. Physical functioning
2. Role-physical
3. Bodily pain
4. General health
5. Vitality
6. Social functioning
7. Role-emotional
8. Mental health

Longitudinal study

- Carried out in Galdakao Hospital (Viscay).
- Goal: measure the health status and evolution of patients being treated for COPD.
- Patients were followed for up to five years.
- 543 patients and 4 measurements.

Longitudinal study

- The 10% of the population in European Union suffers this disease.
- It is characterized by time to time deterioration.
- It is the fourth cause of death due to disease in the world.
- There is a large variability in the medical decision procccess, including hospitalization criterion.

Longitudinal study

- Carried out in Galdakao Hospital (Viscay).
- Goal: measure the health status and evolution of patients being treated for COPD.
- Patients were followed for up to five years.
- 543 patients and 4 measurements.

Literature

Beta-binomial distribution for analysing HRQoL (Arostegui et al., 2007).

Beta-binomial regression:

- "logistic" regression (Forcina and Franconi 1988).
- Hierarchical GLM (Lee and Nelder 1996).

Literature

Beta-binomial distribution for analysing HRQoL (Arostegui et al., 2007).

Beta-binomial regression:

- "logistic" regression (Forcina and Franconi 1988).
- Hierarchical GLM (Lee and Nelder 1996).

"logistic" regression is more adequate in HRQoL data (Najera-Zuloaga et al., 2016).

Problem/Theoretical motivation

Beta-binomial distribution:

- Consists of a sum of correlated Bernoulli variables.

$$y \sim \text{BB}(m, p, \phi) \Leftrightarrow y|u \sim \text{Bin}(m, u), \;\; u \sim \text{Beta}(p/\phi, (1-p)/\phi)$$

$$\text{E}[y] = mp, \qquad \text{Var}[y] = mp(1-p)[1 + (n-1)\frac{\phi}{1+\phi}]$$

Problem/Theoretical motivation

Beta-binomial distribution:

- Consists of a sum of correlated Bernoulli variables.

$$y \sim \text{BB}(m, p, \phi) \Leftrightarrow y|u \sim \text{Bin}(m, u), \;\; u \sim \text{Beta}(p/\phi, (1-p)/\phi)$$

$$\text{E}[y] = mp, \qquad \text{Var}[y] = mp(1-p)[1 + (n-1)\frac{\phi}{1+\phi}]$$

Problem:

- Beta-binomial does not belong to exponential family.
- No mixed model developed:

$$\eta = \text{logit}(p/(1-p)) = X\beta + Zu$$

Definition

Given some outcome variables $y_1, y_2, \ldots, y_n$, we have that

$$y_i|u \sim \text{BB}(m, p_i, \phi) \qquad \text{and} \qquad u \sim \mathcal{N}(0, D).$$

Definition

Given some outcome variables $y_1, y_2, \ldots, y_n$, we have that

$$y_i|u \sim \text{BB}(m, p_i, \phi) \qquad \text{and} \qquad u \sim \mathcal{N}(0, D).$$

Given the covariates $X_1, X_2, \ldots, X_p$, we construct the following linear predictor:

$$\eta_i = \text{logit}(p_i) = x_i'\beta + z_i'u$$

where x_i and z_i are the ith row of full rank matrices X and Z, consisting of the given covariates and the random structure of the model.

Estimation process

It is focused on the following approximated likelihood

$$l(\beta, \theta | y, \tilde{u}) = -\frac{1}{2}\log(\det(D^*)) + h(\tilde{u})$$

where,

h is the hierarchical likelihood defined by Lee and Nelder (1996),

$\tilde{u}$ is the solution of $\frac{\partial h}{\partial u} = 0$,

and, $D^* = \frac{\partial^2 h}{\partial u \partial u'}|_{u=\tilde{u}}$.

Creating the data

Simulated longitudinal scenario:

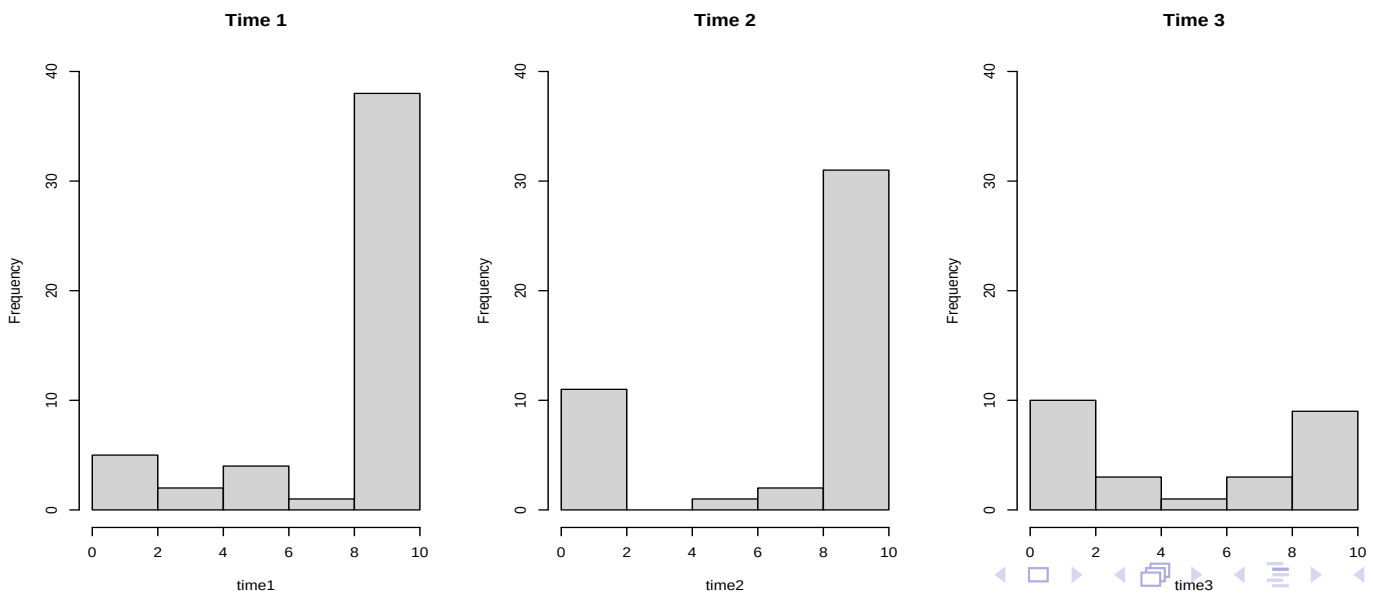
- n^o of individuals 50, n^o of time measurements 3.
- $x \sim \mathcal{N}(1, 1)$.
- $\beta_0 = -1, \beta_1 = 2$.
- $u \sim \mathcal{N}(0, \sigma^2 I_3)$, where $\sigma = 2$.
- $\eta_i = x_i'\beta + z_i'u \Longrightarrow p_i = 1/(1 + exp(-\eta_i))$.
- $\phi = 2$ and $m = 10$.
- $y_i \sim \text{BB}(m, p_i, \phi)$.

The toy data

ID	$t = 1$	$t = 2$	$t = 3$
1	10	10	10
2	10	2	NA
3	10	NA	NA
⋮	⋮	⋮	⋮
50	1	10	NA
	$n_1 = 50$	$n_2 = 45$	$n_3 = 26$

The toy data

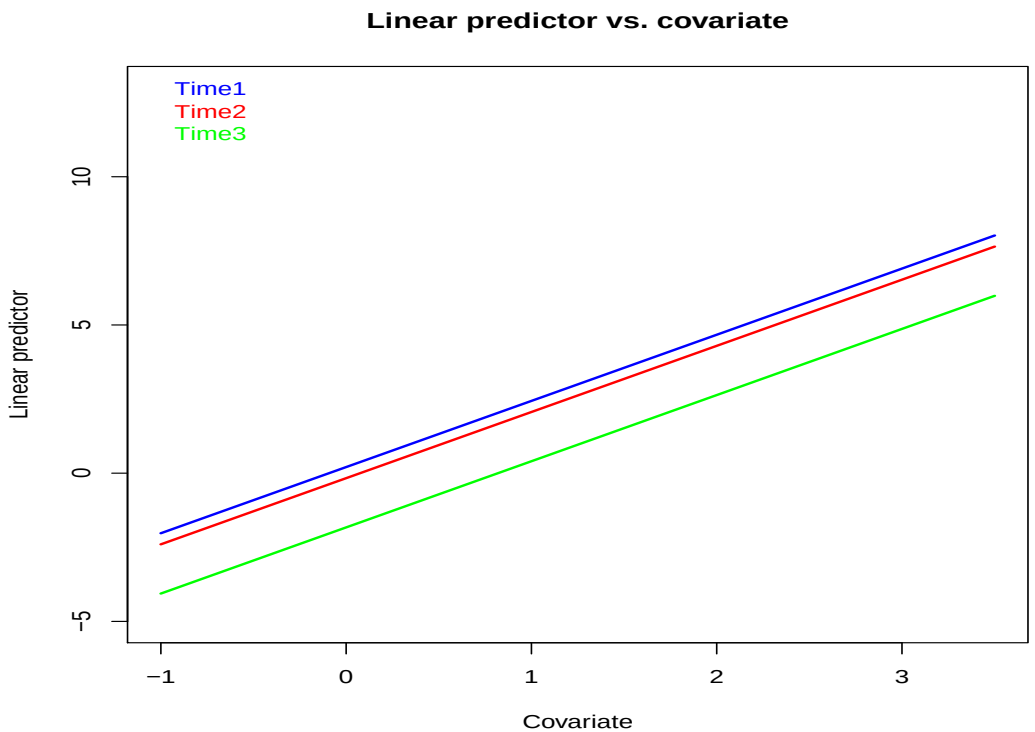
ID	$t = 1$	$t = 2$	$t = 3$
1	10	10	10
2	10	2	NA
3	10	NA	NA
⋮	⋮	⋮	⋮
50	1	10	NA
	$n_1 = 50$	$n_2 = 45$	$n_3 = 26$



Different models

$$\boxed{\Delta x \rightarrow \Delta y_t}$$

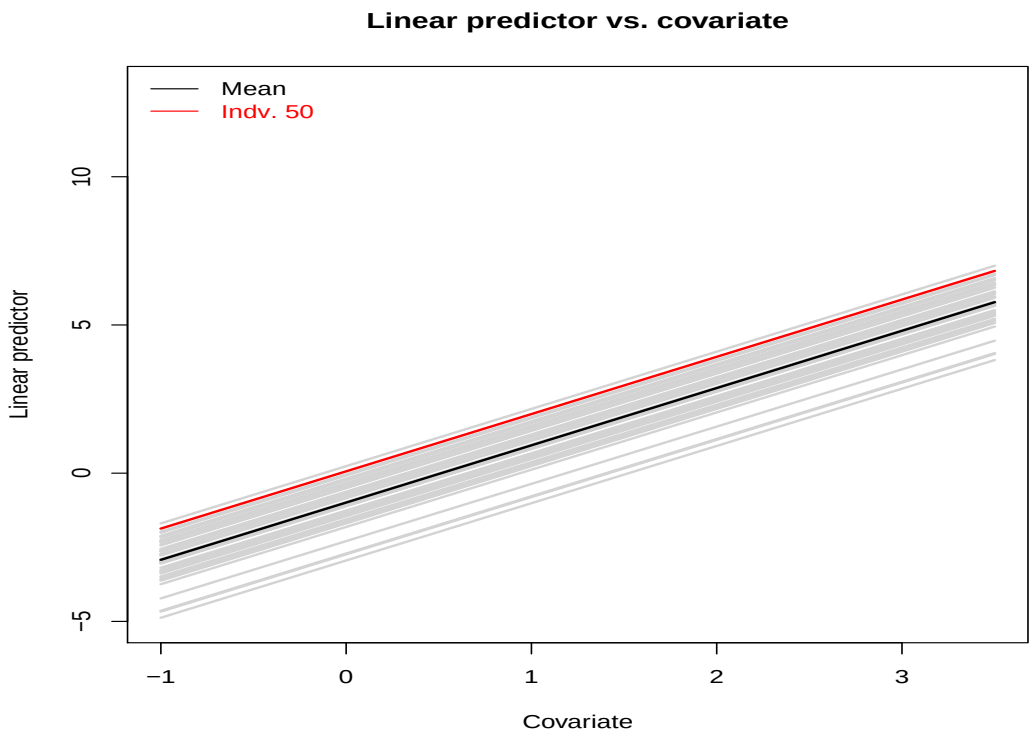
Random intercept per time: $\eta_{it} = \beta_0 + x_{it}\beta_1 + u_t$



Different models

$$\boxed{\Delta x \rightarrow \Delta y_i}$$

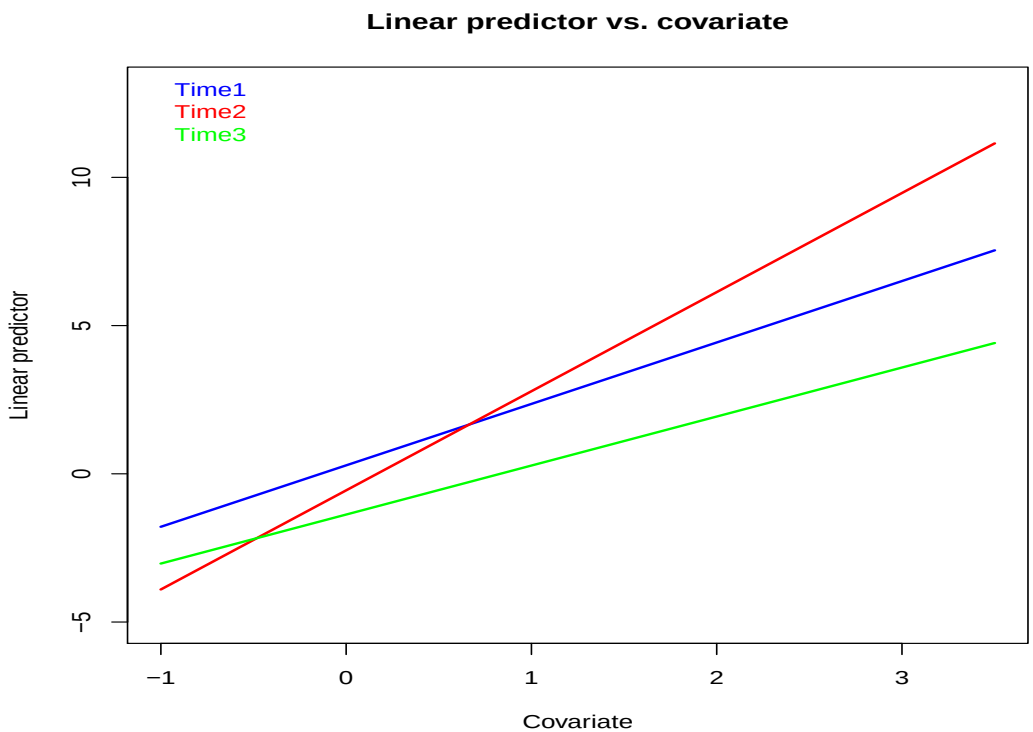
Random intercept per individual: $\eta_{it} = \beta_0 + x_{it}\beta_1 + v_i$



Different models

$$\boxed{\Delta x \rightarrow \Delta y_t}$$

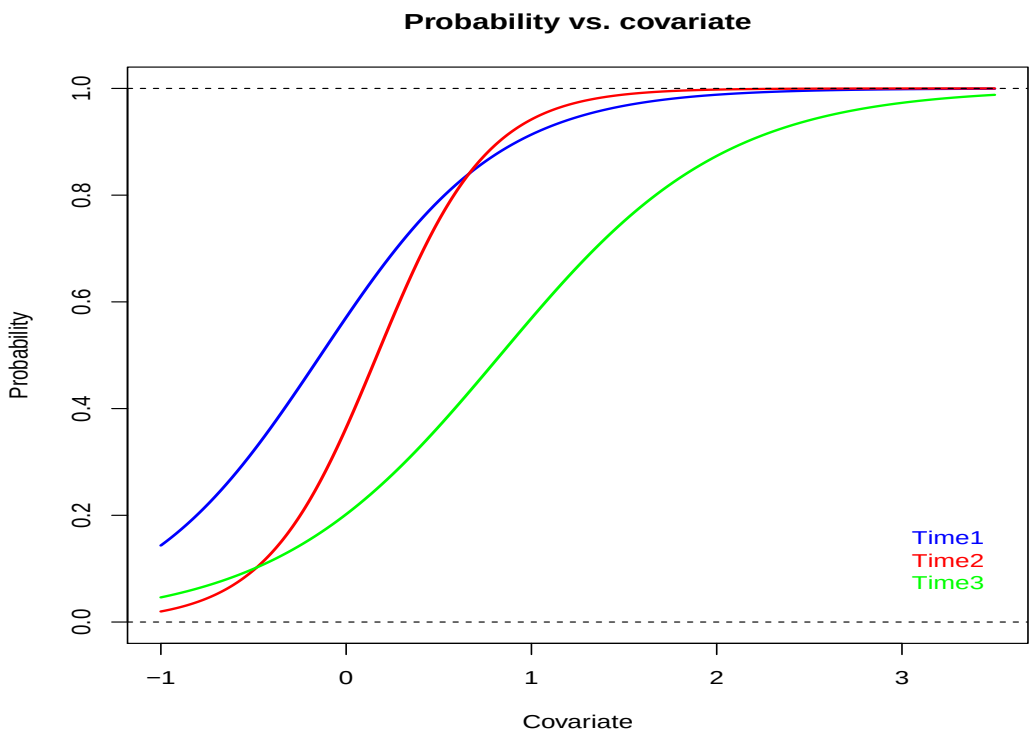
Random intercept and slope: $\eta_{it} = \beta_0 + x_{it}(\beta_1 + \beta_t) + u_i + v_t$



Different models

$$\boxed{\Delta x \rightarrow \Delta y_t}$$

Random intercept and slope: $\eta_{it} = \beta_0 + x_{it}(\beta_1 + \beta_t) + u_i + v_t$



Package example

`R`-package: `HRQoL`

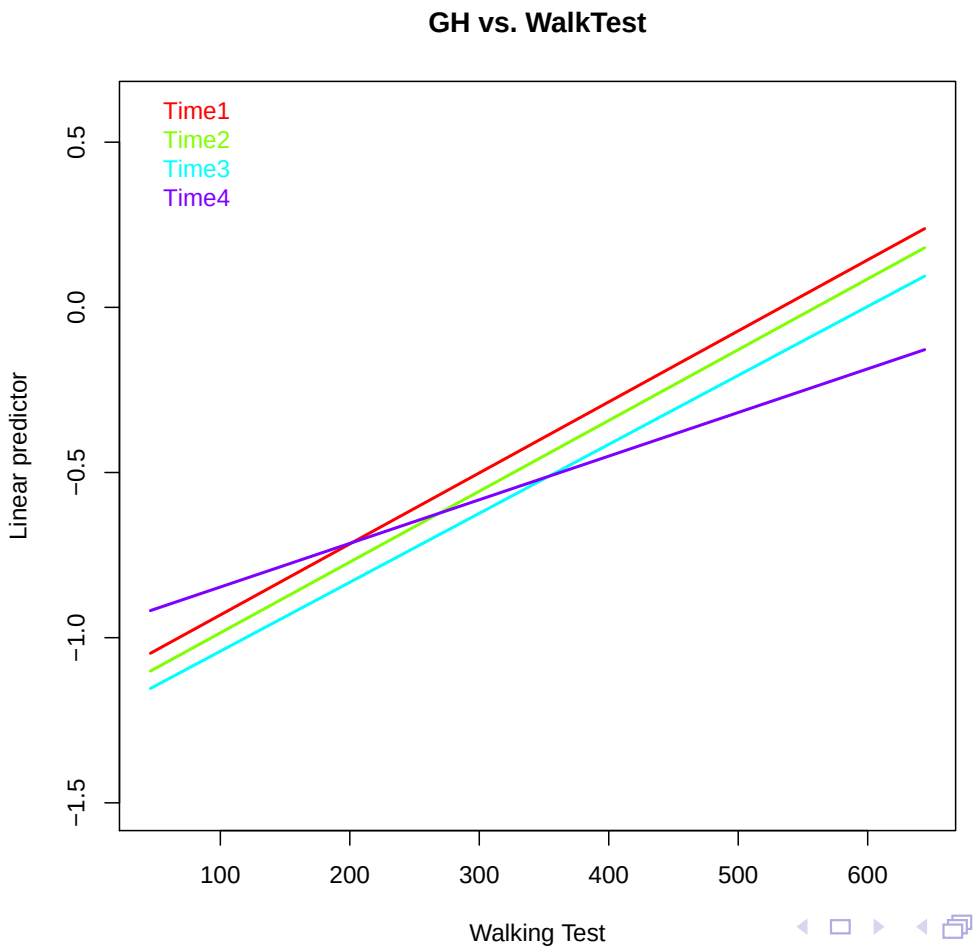
- HRQoL plots.
- Recodes.
- Beta-binomial distribution based analysis.
- Longitudinal models.
- ....

R-package: HRQoL

```
Call: BBmm(fixed.formula = y~x, random.formula=~ID+time, m, data)

Fixed effects coefficients:

            Estimate   StdErr t.value   p.value
(Intercept) 1.001291 0.096883  10.335 < 2.2e-16 ***
x           2.005359 0.101176  19.820 < 2.2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
-------------------------------------------------------------

Random effects standard error coefficients:

     Estimate StdError t.value  p.value
[1,]  0.95878  0.26539  3.6127 0.000303 ***
[2,]  1.01272  0.31701  3.1946 0.001400 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Number of random components in each random effect: 11 6
-------------------------------------------------------------

Beta-binomial transformed dispersion parameter coefficients:

    Estimate StdError t.value   p.value
psi -1.06365  0.11366 -9.3579 < 2.2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
-------------------------------------------------------------

Number of observations: 1000

Number of iterations: 5
```

Real application: General health dimension

$$\eta_{it} = \beta_0 + (\beta_1 + \beta_t)\text{WalkingTest}_{ij} + u_i + v_t$$



Future work

- Structure of the random effects: Temporal correlation.
- Different variances.
- Multidimensional framework: All dimensions together.

Bibliography

[1] I. Arostegui, V. Núñez Antón, and J. M. Quintana. Analysis of short form-36 (sf-36): The beta-binomial distribution approach. *Statistics in Medicine*, 26:1318–1342, 2006.

[2] I. Arostegui, V. Núñez Antón, and J. M. Quintana. Statistical approaches to analyse patient-reported outcome as response variables: An application to health-related quality of life. *Statistical Methods in Medical Research*, 21(2):189–214, 2010.

[3] I. Arostegui, V. Núñez Antón, and J. M. Quintana. On the recoding of continuous and bounded indexes to a binomial form: an application to quality-of-life scores. *Journal of Applied Statistics*, 40(3):563–582, 2013.

[4] N. E. Breslow and D. G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25, Mar. 1993.

[5] A. Forcina and L. Franconi. Regression analysis with the beta-binomial distribution. *Rivista di Statistica Applicata*, 21(1), 1988.

[6] Y. Lee and J. A. Nelder. Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(4):619–678, 1996.

[7] Y. Lee, J. A. Nelder, and P. Yudi. *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood*. Chapman & Hall/CRC, 2006.

[8] J. Najera-Zuloaga, D.-J. Lee, and I. Arostegui. Comparison of beta-binomial regression model approaches to analyze health realted quality of life data. *Under review*.

[9] Y. Pawitan. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press, 2001.

[10] R. Schall. Estimation in generalized linear models with random effects. *Biometrika*, 78(4):719–27, 1991.

ESKERRIK ASKO

GRACIAS