Comparison of beta-binomial regression model approaches to analyze health related quality of life data

January 4, 2017

Josu Najera-Zuloaga [1], Dae-Jin Lee[1] and Inmaculada Arostegui[1,2,3]

[1]Basque Center for Applied Mathematics, 48009 Bilbao, Spain.
[2]Universidad del País Vasco UPV/EHU. Departamento de Matemática Aplicada y Estadística e Investigación Operativa. Aptdo. de correos 644. 48080 Bilbao, Spain.
[3]Red de Investigación en Servicios de Salud en Enfermedades Crónicas - REDISSEC

Abstract

Health related quality of life (HRQoL) has become an increasingly important indicator of health status in clinical trials and epidemiological research. Moreover, the study of the relationship of HRQoL with patients' and disease's characteristics has become one of the primary aims of many HRQoL studies. HRQoL scores are usually assumed to be distributed as binomial random variables and often highly skewed. The use of the beta-binomial distribution in the regression context has been proposed to model such data, however, the beta-binomial regression has been performed by means of two different approaches in the literature: i) beta-binomial distribution with a logistic link; and ii) hierarchical generalized linear models (HGLMs). None of the existing literature in the analysis of HRQoL survey data has performed a comparison of both approaches in terms of adequacy and regression parameter interpretation context.

This paper is motivated by the analysis of a real data application of HRQoL outcomes in patients with Chronic Obstructive Pulmonary Disease (COPD), where the use of both approaches yields to contradictory results in terms of covariate effects significance and consequently the interpretation of the most relevant factors in HRQoL. We present an explanation of the results in both methodologies through a simulation study and address the need to apply the proper approach in the analysis of HRQoL survey data for practitioners, providing an `R` package.

KEYWORDS: Beta-binomial regression, Hierarchical GLM, Health Related Quality of Life, Chronic Obstructive Pulmonary Disease, `HRQoL R` package.

1 Introduction

It has previously stated that clinical and laboratory outcomes are not necessarily the most important outcome results in studies that examine the effect of health interventions (Testa and Simonson, 1996). Moreover, patient-reported outcomes are increasingly used as primary outcome measures in observational and experimental studies. For instance, health related quality of life (HRQoL) has become an important outcome as a measure of health status and a way to evaluate medical care results, especially in chronic diseases. The measurement of the HRQoL provides information about the disease and its impact in the patient in a standardized, comparable and objective way (Goldsmith, 1972). Measuring HRQoL can help determine the burden of preventable disease, injuries, and disabilities, and it can provide valuable new insights into the relationships between HRQoL and risk factors. Therefore, to study the relationship of HRQoL with patients' and disease's characteristics has become one of the primary aims of many HRQoL studies.

Several instruments for measuring HRQoL have been developed in form of questionnaires, some of them in a generic way and other for specific diseases. One of the most widely used generic instruments is the Short Form-36 (SF-36) Health Survey. The SF-36 has 36 items that are reduced to 8 health dimensions, it was developed and validated and it has been translated and validated into many languagesWare et al. (1993).

The beta-binomial distribution has been proposed in the literature to fit HRQoL data provided by the SF-36. In this context, a regression model based on the beta-binomial distribution has been shown to be a good option, not only to detect significant predictors of HRQoL when SF-36 is used, but also to analyze and interpret the effect of several explanatory variables on HRQoL Arostegui et al. (2007). The comparison of this regression method based on the beta-binomial distribution with other commonly used modelling approaches in the same context (such as probit and multiple linear regression) showed that the beta-binomial distribution was a satisfactory approach to account for over-dispersion Arostegui et al. (2012). However, there are two different ways of implementing a regression model based on the beta-binomial distribution in the literature. Firstly, a regression approach based on the use of a logistic link function in the probability parameter of a beta-binomial distribution (BB logistic linkForcina and Franconi (1988)). In this setting, estimation is done via maximum likelihood and iterative re-weighted least squares in a Generalized Linear Model (GLM) fashion. On the other hand, the Hierarchical Generalized Linear Model (HGLMLee and Nelder (1996)) as a generic method of performing Generalized Linear Mixed Models (GLMMs) with non-normal random effects, which allows the inclusion of beta distributed random effects in the linear predictor of the probability parameter of a binomial distribution.

There are some references in the literature that compare the maximum likelihood estimation approach, which the BB logistic link is based on, and the HGLM approach for estimating parameters of the beta-binomial distribution, concluding that both methodologies are valid and equivalent to fit beta-binomial distribution to some data if some second order approximations are used in HGLM (Lee et al., 2007; Lee and Lee., 2012). However, although it has been stated that the beta-binomial regression is a good option to analyze HRQoL data, none comparison between these two different approaches has been addressed from a practical point of view in a regression context where the interest lies in the estimation and interpretation of covariate effects.

The objective of this work is to show that the different beta-binomial regression approaches lead to contradictory results, both in regression parameters estimation and statistical significance tests, and to provide the best option to analyze HRQoL data from the SF-36 Health Survey. In fact, the motivation of this paper was the application of both approaches to real HRQoL data in patients with stable Chronic Obstructive Pulmonary Disease (COPD) which yielded to different estimates of the regression coefficients misleading the interpretation of the covariate effects in

some circumstances. A simulation study was performed to check the different results provided by both approaches under several controlled scenarios.

The rest of the paper is organized as follows. Section 2 presents a description of both methodological approaches to perform a beta-binomial regression analysis in depth, together with the widely used mixed effects logistic regression model. The application of both beta-binomial based methodologies to real data is presented in Section 3, starting with a description of the COPD study, continuing with the description and recoding of the SF-36 Health Survey and finishing with the comparison of the results obtained after applying both methodologies. Section 4 is focused in the simulation study that provides comparisons of all approaches in controlled scenarios. Finally, in Section 5 we provide a brief discussion of the results obtained, as well as some general conclusions and recommendations.

2   Alternative approaches for fitting beta-binomial regression

In this Section, we introduce the beta-binomial distribution and the notation we will follow through the rest of the paper.

The beta-binomial distribution consists of a finite sum of Bernoulli dependent variables whose probability parameter is random and follows a beta distribution. Assume that we have y_j a set of variables, $j = 1, \ldots, m$ with $m \in \mathbb{N}$, that conditioned on a random variable u, are independent and follow a Bernoulli distribution with parameter u. On the other hand, the random variable u, which we will denote the probability variable, follows a beta distribution with parameters α_1 and α_2. Namely,

$$y_j | u \sim \text{Ber}(u) \qquad \text{and} \qquad u \sim \text{Beta}(\alpha_1, \alpha_2)$$

with, $\text{E}[u] = \psi$ and $\text{Var}[u] = \psi(1 - \psi)\phi/(1 + \phi)$, where $\psi = \alpha_1/(\alpha_1 + \alpha_2)$ and $\phi = 1/(\alpha_1 + \alpha_2)$. Consequently, the marginal moments of first and second order of the variables are

$$\text{E}[y_j] = \psi \qquad \text{and} \qquad \text{Var}[y_j] = \psi(1 - \psi),$$

and, ρ, the intraclass correlation is

$$\rho = \text{Corr}[y_j, y_k] = \frac{\phi}{1 + \phi}, \quad j, k = 1, \ldots, m, \text{ where } j \neq k, \tag{1}$$

which determinates the parameter ϕ as the correlation parameter.

If we sum up all the variables we will define a new variable,

$$y = \sum_{j=1}^{m} y_j$$

which follows a new distribution that is called beta-binomial distribution, and it is defined as follows. The variable y follows a beta-binomial distribution if

$$y | u \sim \text{Bin}(m, u) \qquad \text{and} \qquad u \sim \text{Beta}(\alpha_1, \alpha_2). \tag{2}$$

The marginal density of the response variable y can be written explicitly as

$$f(y) = \int_0^1 f_{y|u}(y|u) f_u(u) du \tag{3}$$

$$= \binom{m}{y} \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1 + \alpha_2 + m)} \frac{\Gamma(\alpha_1 + y)}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_2 + m - y)}{\Gamma(\alpha_2)}.$$

3

2.1 Beta-binomial logistic regression

First, we consider the beta-binomial distribution of maximum number of scores m as a given distribution of a set of response variables y_i, $i = 1, \ldots, n$. If we perform a reparameterization proposed by Arostegui et al. Arostegui et al. (2007) in the beta-binomial model parameters considering $\alpha_{1i} = p_i/\phi$ and $\alpha_{2i} = (1 - p_i)/\phi$, we have

$$\mathrm{E}[y_i] = mp_i \qquad \text{and} \qquad \mathrm{Var}[y_i] = mp_i(1 - p_i)\left[1 + (m - 1)\frac{\phi}{1 + \phi}\right]. \tag{4}$$

Hence, given equation (4), we can interpret the beta-binomial distribution as a binomial distribution with some over-dispersion structure which is given by the intraclass correlation through the summed Bernoulli observations. Therefore, similarly to the binomial case, p_i is interpreted as the probability of success in each Bernoulli observation of individual i. Notice that when $\phi = 0$, the model corresponds with the binomial case.

If we use the Gamma function property, $\Gamma(x + 1) = x\Gamma(x)$ for all $x \in \mathbb{R}$ and $\Gamma(0) = 1$, we get a new expression of the beta-binomial density function in equation (3), which is easier to manipulate. Hence, the log-likelihood of $\boldsymbol{y} = (y_1, \ldots, y_n)$ is given by the expression:

$$l(\boldsymbol{p}, \phi | \boldsymbol{y}, m) = \sum_{i=1}^{n} \left[\log \binom{m}{y_i} + \sum_{k=0}^{y_i} \log(p_i + k\phi) + \sum_{k=0}^{m-y_i} \log(1 - p_i + k\phi) \right. $$
$$\left. - \sum_{k=0}^{m} \log(1 + k\phi) \right], \tag{5}$$

where $\boldsymbol{p} = (p_1, \ldots, p_n)$. Using this reparameterization of the beta-binomial distribution parameters, we have considered p_i as the probability of success in each Bernoulli observation of each individual, which allows us to link it with some given covariates by a linear predictor choosing an appropriate link function. For regression models, Forcina et al. Forcina and Franconi (1988) assumed that the probability parameter p_i is connected to a vector of regression parameters by means of a logit link function model, so they formulated the following relationship between the probability parameter of the beta-binomial distribution of each individual and some given covariates $X_1, X_2, \ldots, X_t$ as,

$$p_i = \exp(x_i'\beta)/(1 + \exp(x_i'\beta)), \tag{6}$$

where β is a $(t + 1) \times 1$ vector of regression parameters and x_i', $i = 1, \ldots, n$ where n is the number of observations, is the ith row of a full rank design matrix $\boldsymbol{X}$ composed by the given covariates. By the relationship in equation (6) we obtain the following linear predictor,

$$\eta_i = g(p_i) = \log(p_i/(1 - p_i)) = x_i'\beta. \tag{7}$$

Therefore, replacing equation (6) on equation (5), we can obtain the maximum likelihood estimators of the regression parameters and dispersion parameter ϕ by means of an iterative method based on the maximum likelihood approach to perform estimations and inference in the parameters β and ϕ Forcina and Franconi (1988). We will denote this model as *BBlogit*.

2.2 Hierarchical GLM approach

The generalized linear mixed models (GLMMs) (McCulloch and Searle, 2001), where the linear predictor of a GLM (McCullagh and Nelder, 1989) is allowed to have, in addition to the usual fixed effects, one or more gaussian random effects, is a very widely used methodology in different frameworks. Although the normal distribution is convenient for specifying correlations among

random effects, the use of other distributions for the random effects greatly enriches these class models. Lee and Nelder Lee and Nelder (1996) extended GLMMs to hierarchical GLMs (HGLMs), in which the distribution of random components is extended to conjugates of arbitrary distributions from the GLM family.

The HGLMs are defined as, on the one hand, conditional on some given random components u, the response variable y follows a GLM family, satisfying that

$$\mathrm{E}[y|u] = \mu \qquad \text{and} \qquad \mathrm{Var}[y|u] = \lambda V(\mu)$$

and the linear predictor takes the form

$$\eta = g(\mu) = X\beta + Zv,$$

where $v = v(u)$, the scale, are the random effects and β are the fixed effects. On the other hand, the random components u follow a distribution conjugate to a GLM family of distributions with parameter ϕ.

We have seen in the beginning of Section 2 that, the beta-binomial model consists of a conditioned response following a binomial distribution and some given random components u following a beta distribution. Due to the fact that the beta distribution is conjugate to the binomial GLM, we can consider the beta-binomial model as a special case of the HGLM family, where

$$y_i|u_i \sim \mathrm{Bin}(m, p_i) \qquad \text{and} \qquad u_i \sim \mathrm{Beta}(\alpha_1, \alpha_2)$$

and p_i is connected to u_i by a linear predictor.

When we construct a HGLM we must choose $v(\cdot)$ the scale on which the random effects occur linearly in the linear predictor, that is called *weak canonical scale*. This weak canonical scale allows the model to maintain invariance of inference with respect to equivalent approaches (Lee et al., 2006). Consequently, the linear predictor of the beta-binomial HGLM is

$$\eta_i = \mathrm{logit}(p_i) = x_i'\beta + v_i, \tag{8}$$

where x_i' is the ith row of a full rank design matrix $\mathbf{X}$ composed by the given covariates, β are the fixed effects and $v_i = \mathrm{logit}(u_i)$ is the random effect attributed to individual i, $i = 1, \ldots, n$.

Constraints must be specified in either the random or fixed part of the model to maintain the structure of the model. Lee and Nelder 2001 proposed to impose constraints in the random part of the model, fixing the expectation of the random components u equal to value that the scale transforms to zero. Namely, in the beta-binomial HGLM they imposed,

$$\mathrm{E}[u_i] = 1/2. \tag{9}$$

Consequently, they defined the distribution of the random components u beta with parameters $1/\alpha$, $\alpha > 0$,

$$u_i \sim \mathrm{Beta}(1/\alpha, 1/\alpha), \tag{10}$$

which fixes the previously defined beta distribution parameters $\psi = 1/2$ and $\phi = \alpha/2$.

In HGLM models, especially in the binomial beta HGLM, the computation of the marginal likelihood is not straightforward, and moreover, it is totally uninformative about the random effects v. Consequently, Lee and Nelder Lee and Nelder (1996) proposed the so-called H-likelihood as an approach to perform inference in HGLMs. The H-likelihood is defined by

$$h = h(y, v|\beta, \phi) = \log f(y|v; \beta) + \log f(v; \phi), \tag{11}$$

where the first term is the conditioned log-likelihood of the response, the binomial density function, and the second term is the log-likelihood corresponding to the random effects in the linear predictor, the beta density function through the logit transformation.

For inference in HGLMs three likelihoods are available, the H-likelihood and two adjusted profile likelihoods, namely the marginal likelihood $\mathcal{L}$, that eliminates nuisance random effects v from h by integration, and the restricted likelihood, that eliminates nuisance fixed effects β from $\mathcal{L}$ by conditioning on the maximum likelihood estimates of β. In principle we should use the H-likelihood h for inference about v, the marginal-likelihood $\mathcal{L}$ for β, and the restricted likelihood for the dispersion parameters (Lee et al., 2006).

2.3   Mixed effects logistic regression model

The mixed effects logistic regression model is a special case of the GLMMs, which have been briefly introduced in the beginning of Section 2.2. First of all, we have to clarify that this methodology is not based on the beta-binomial distribution. Consequently, due to the nature of the paper, we have not included it in the real application section to avoid confusion to the readers. However, simulations based on this approach have been performed and included in Section 4, as it has been previously mentioned, the GLMM is a very widely used methodology even in HRQoL framework. That way we can perform a full simulation study, not only concluding which beta-binomial based approach is more appropriate for analyzing HRQoL, but also comparing them with a commonly used methodology. For the sake of clarity, we will explain briefly in this section what the mixed effect logistic regression consists of.

In the special case of the mixed effects logistic regression model, we assume that the responses y_i, $i = 1, \ldots, n$, given some gaussian distributed random effects u, are conditionally independent and follow a binomial distribution. Namely,

$$y|u \sim \mathrm{Bin}(m, p_i) \qquad u \sim \mathcal{N}(0, D).$$

For analyzing the HRQoL, due to the variability that exists between different individuals, we are going to assume that each individual is connected with a component of the random effects u. Consequently, we are going to assume that the number of random effects is equal to the number of individuals, that there is no correlation between the individuals and that the variance remains constant within individuals. Namely, the variance-covariance matrix of the random effects is

$$D = \sigma_u^2 I_n.$$

The model defines a linear predictor η_i by assuming the following relationship between some given covariates $X_1, \ldots, X_t$ and the probability parameter of each conditioned response variable y_i,

$$\eta_i = g(p_i) = x_i'\beta + u_i, \qquad i = 1, \ldots, n,$$

where $g(\cdot)$ is the logistic link function defined in equation (7), β is a $(t+1) \times 1$ vector of fixed regression parameters, x_i' is the ith row of a full rank design matrix $\mathbf{X}$ composed by the given covariates and u_i is the random effect connected with ith individual.

The basis of the likelihood approximation in GLMM is the extended likelihood that includes the random effects, plus the heuristics provided by the normal mixed models,

$$\mathrm{log}L(\beta, \sigma_u, u) = \mathrm{log}f_{y|u}(y|u) + \mathrm{log}f_u(u)$$

where $f_{y|u}(y|u)$ is the conditional density function of the responses, in the mixed effects logistic regression model the binomial density function, and $f_u(u)$ is the gaussian density function of the random effects. Schall Schall (1991) provided an easy algorithm for performing the estimation based on the extended likelihood approach and Breslow and Clayton Breslow and Clayton (1993) presented applications that support its practical viability.

3 Application to real data: COPD Study

3.1 Description of the study

Chronic Obstructive Pulmonary Disease is a very common chronic disease around the world. Some well-designed studies have found a measured prevalence of COPD in Europe between 4% and 10% of adults and it is expected to increase over the next years Halbert et al. (2003); Buist et al. (2008). According to estimates of the World Health Organization (WHO), by 2020 it shall become the third most frequent cause of death, following coronary and cerebrovascular diseases Murray and López (1997). Moreover, patients with COPD generally have a substantial worsening in their quality of life Esteban et al. (2009).

Researchers of the Respiratory Service at Hospital Galdakao in Spain designed the COPD Study, a longitudinal study whose main goal was to measure the health status and evolution of patients being treated for COPD. Patients were recruited at five outpatient respiratory clinics affiliated with the hospital and consecutively included in the study for one year, starting at January 2003. Patients were eligible for the study if they had been diagnosed with COPD for at least six months and they had been receiving medical care at one of the hospital respiratory outpatient facilities for at least six months. Their COPD had to be stable for six weeks before enrolment. Patients were followed for up to five years. One of the aims of the study was to measure the effect clinical and socio-demographic variables have on the HRQoL for patients with stable COPD. In order to do that, data were selected from the patients' first visit to the COPD study, including a sample of 543 patients. A set of selected variables recoded in the study and included as covariates in the models were socio-demographic variables such as gender and age, together with forced expiratory volume in one second in percentile (FEV1%), body mass index (BMI), dyspnea (measured with the modified scale of the Medical Research Council Mahler et al. (2009)), the 6-minute walking testsAmerican Thoracic Society (2002) and presence of anxiety and depression measured by the Hospital Anxiety and Depression scale (HADZigmond and Snaith (1983)) among others. A socio-demographic and clinical summary of the exploratory variables is presented in Table 1.

[Table 1 about here.]

3.2 Description of the SF-36 Health Survey

The SF-36 Health Survey was developed within the Medical Outcomes Study (Ware et al., 1993). It measures generic HRQoL concepts and it provides an objective way to measure HRQoL from the patients point of view by scoring standardized responses to standardized questions. The validity and reliability of this instrument has been broadly tested (Stansfeld et al., 1997). The SF-36 questionnaire has 36 items, with different answer options. It was constructed to represent eight health dimensions, which are *physical functioning* (PF), *role physical* (RP), *body pain* (BP), *general health* (GH), *vitality* (VT), *social functioning* (SF), *role emotional* (RE) and *mental health* (MH). Each item is assigned to a unique health dimension. Each of the eight multi-item dimensions contains two to ten items. The first four dimensions are mainly physical, whereas the last four measure mental aspects of HRQoL. The standardized scoring system is thoroughly described by the original authors (Ware et al., 1993). Briefly, each score was calculated with an algorithm based on the original items assigned to this dimension. For each dimension, the answers to the items were first recoded and then added in a weighted sum fashion. The resulting raw scores were then transformed to standardized scale scores from 0 to 100, where a higher score indicates a better health status.

Consequently, the SF-36 Health Survey generates a profile of HRQoL outcomes on eight health dimensions. Additionally, the SF-36 includes an item related to health transition, which is not used in the scoring of the eight health dimensions. The authors of the SF-36 Health

Survey also provide normative scores for each health dimension. Each SF-36 score was first standardized using the mean and standard deviations obtained from the general population and then transformed to a norm-based (mean=50, standard deviation = 10) scoring (SF-36 Health Survey, 1993). Two summary measures, one physical and one mental, can be created from the eight main domains. These two summary scores were generated using the physical and mental factor score coefficients from the general population and they were also transformed to norm-based scoring (Ware et al., 1994). However, we have used neither norm-based scores nor summary scores in this work.

The SF-36 Health Survey was rated in 2002 by the British Medical Journal (Garratt et al., 2002) as the most frequently used patient reported outcome of generic health in the scientific publications. In this work, generic HRQoL was measured using version 1.2 of the SF-36 Health Survey, which corresponds to the version 1.4 of the Spanish version.

In order to provide a better understanding of the construction of the SF-36 Health Survey dimensions and COPD data, we show in Table 4 the number of items related to each dimension and the number of possible values each dimension can obtain, together with the mean and standard deviation of the original standardized SF-36 dimensions in the COPD data. In addition, we provide in Appendix A Figure 8 where the distribution of the original standardized SF-36 dimensions in COPD data is shown. In Figure 8 the different shapes of the original standardized SF-36 scores can be appreciated. Moreover, in Appendix A, Table 12 shows the mean and the standard deviation of the SF-36 original standardized scores stratified by categorical covariates of interest.

[Table 2 about here.]

3.3   Recoding of the SF-36 Health Survey scores to a binomial form

Within this section, we will briefly describe the process of analyzing HRQoL data from the SF-36 Health Survey using the beta-binomial regression approach. It includes the proposal, validation, comparison to other methods of analysis and the recoding approach necessary in order to fit the beta-binomial model. All the process has been thoroughly described previously. What we present here is only a summary that could help readers to contextualize the results we show in the present work.

The beta-binomial distribution was proposed in 2007 to fit the SF-36 Health Survey scores (Arostegui et al., 2007). The proposal was mainly motivated because of the ordinal feature that many of the HRQoL scales exhibit. They showed that the beta-binomial regression was a good option to detect significant predictors of HRQoL and they also provided a nice interpretation of the effect of explanatory variables on HRQoL when SF-36 is used. The authors also compare results using multiple linear regression (MLR) and beta-binomial regression for real and simulated data, showing that performance of the beta-binomial approach was better or similar than the MLR approach in all the HRQoL dimensions of the SF-36. Comparison of MLR and beta-binomial regression approaches was performed based on distributional assumptions. The paper closed recommending the analysis of all the HRQoL scores of the SF-36 with the same method, consequently, the authors recommended to use the beta-binomial regression.

A posterior and more general publication of the same authors (Arostegui et al., 2012) presented eight methods of analysis of patient reported outcomes, such as HRQoL, under different assumptions that lead to different interpretations of the results. The methods were: MLR, with least square and bootstrap estimations, tobit regression, ordinal logistic and probit regressions, beta-binomial regression, binomial-logit-normal regression (BLNR) and coarsening. All methods were applied to scores obtained from two of the health dimensions of the SF36 Health Survey. The authors showed that the beta-binomial regression approach renders satisfactory results in a broad number of situations, with a very convenient clinical interpretation of the results.

8

Finally, recoding of the scores to a binomial form is necessary in order to fit the beta-binomial distribution to the SF-36 Health Survey scores. The same authors proposed and evaluated a method of recoding continuous and bounded scores, such as HRQoL scores, to a binomial form (Arostegui et al., 2013). The method was mainly based on the possible number of values each dimension can obtain, which, as it has been explained in Section 3.1 and showed in Table 4, comes from the number of items related with the construction of each dimension. Indeed, the methodology transforms the sorted scale of possible values of each dimension score to a ordinal scale from 0 to m, i.e. to a binomial form scale. The real interval [0,100], which is the scale of the standardized original scores, is divided in some subintervals, and then, each subinterval is linked to the value that corresponds in the order of the $0 - m$ scale, where $m + 1$ is the number of intervals. Consequently, score values within each subinterval are recoded with the value the subinterval was linked in the $0 - m$ scale. The way the subintervals are constructed is the main contribution of Arostegui et al. (2013)Arostegui et al. (2013) and the subdivision of the $0 - 100$ scale for each dimension is available in the Appendix of the mentioned work. We present in Table 5 the recoding process, step by step, for the dimensions that we will analyze in Section 3.6. Table 5 shows the possible values of the raw and standardized scores, as well as the subinterval division of the standardized original scores and the final recoded scores for *physical functioning*, *mental health* and *role emotional* dimensions. The authors evaluated and validated the proposed method of recoding with the scores provided by the SF-36 Health Survey. They showed that the recoding has a natural interpretation, not only for ordinal scores, but also for questionnaires with many dimensions and different profiles, where a common method of analysis is desired, such as the SF-36. Briefly, let Y denote the original standardized score observed in $[0, 100]$ and Y^* the recoded ordinal score, from 0 to m, where $Y^* \sim \text{Bin}(m, p)$. Thus, Y^* could be interpreted as grouped data for a dichotomous outcome that represents the number of successes in m binomial trials and p represents the probability of success in each trial. In the HRQoL context, Y^* is interpreted as the number of "points" that an individual has and p as the probability of obtaining one point more.

[Table 3 about here.]

3.4 Description of the SF-36 Health Survey

The SF-36 Health Survey was developed within the Medical Outcomes Study (Ware et al., 1993). It measures generic HRQoL concepts and it provides an objective way to measure HRQoL from the patients point of view by scoring standardized responses to standardized questions. The validity and reliability of this instrument has been broadly tested (Stansfeld et al., 1997). The SF-36 questionnaire has 36 items, with different answer options. It was constructed to represent eight health dimensions, which are *physical functioning* (PF), *role physical* (RP), *body pain* (BP), *general health* (GH), *vitality* (VT), *social functioning* (SF), *role emotional* (RE) and *mental health* (MH). Each item is assigned to a unique health dimension. Each of the eight multi-item dimensions contains two to ten items. The first four dimensions are mainly physical, whereas the last four measure mental aspects of HRQoL. The standardized scoring system is thoroughly described by the original authors (Ware et al., 1993). Briefly, each score was calculated with an algorithm based on the original items assigned to this dimension. For each dimension, the answers to the items were first recoded and then added in a weighted sum fashion. The resulting raw scores were then transformed to standardized scale scores from 0 to 100, where a higher score indicates a better health status.

Consequently, the SF-36 Health Survey generates a profile of HRQoL outcomes on eight health dimensions. Additionally, the SF-36 includes an item related to health transition, which is not used in the scoring of the eight health dimensions. The authors of the SF-36 Health Survey also provide normative scores for each health dimension. Each SF-36 score was first

standardized using the mean and standard deviations obtained from the general population and then transformed to a norm-based (mean=50, standard deviation = 10) scoring (SF-36 Health Survey, 1993). Two summary measures, one physical and one mental, can be created from the eight main domains. These two summary scores were generated using the physical and mental factor score coefficients from the general population and they were also transformed to norm-based scoring (Ware et al., 1994). However, we have used neither norm-based scores nor summary scores in this work.

The SF-36 Health Survey was rated in 2002 by the British Medical Journal (Garratt et al., 2002) as the most frequently used patient reported outcome of generic health in the scientific publications. In this work, generic HRQoL was measured using version 1.2 of the SF-36 Health Survey, which corresponds to the version 1.4 of the Spanish version.

In order to provide a better understanding of the construction of the SF-36 Health Survey dimensions and COPD data, we show in Table 4 the number of items related to each dimension and the number of possible values each dimension can obtain, together with the mean and standard deviation of the original standardized SF-36 dimensions in the COPD data. In addition, we provide in Appendix A Figure 8 where the distribution of the original standardized SF-36 dimensions in COPD data is shown. In Figure 8 the different shapes of the original standardized SF-36 scores can be appreciated. Moreover, in Appendix A, Table 12 shows the mean and the standard deviation of the SF-36 original standardized scores stratified by categorical covariates of interest.

[Table 4 about here.]

3.5 Recoding of the SF-36 Health Survey scores to a binomial form

Within this section, we will briefly describe the process of analyzing HRQoL data from the SF-36 Health Survey using the beta-binomial regression approach. It includes the proposal, validation, comparison to other methods of analysis and the recoding approach necessary in order to fit the beta-binomial model. All the process has been thoroughly described previously. What we present here is only a summary that could help readers to contextualize the results we show in the present work.

The beta-binomial distribution was proposed in 2007 to fit the SF-36 Health Survey scores (Arostegui et al., 2007). The proposal was mainly motivated because of the ordinal feature that many of the HRQoL scales exhibit. They showed that the beta-binomial regression was a good option to detect significant predictors of HRQoL and they also provided a nice interpretation of the effect of explanatory variables on HRQoL when SF-36 is used. The authors also compare results using multiple linear regression (MLR) and beta-binomial regression for real and simulated data, showing that performance of the beta-binomial approach was better or similar than the MLR approach in all the HRQoL dimensions of the SF-36. Comparison of MLR and beta-binomial regression approaches was performed based on distributional assumptions. The paper closed recommending the analysis of all the HRQoL scores of the SF-36 with the same method, consequently, the authors recommended to use the beta-binomial regression.

A posterior and more general publication of the same authors (Arostegui et al., 2012) presented eight methods of analysis of patient reported outcomes, such as HRQoL, under different assumptions that lead to different interpretations of the results. The methods were: MLR, with least square and bootstrap estimations, tobit regression, ordinal logistic and probit regressions, beta-binomial regression, binomial-logit-normal regression (BLNR) and coarsening. All methods were applied to scores obtained from two of the health dimensions of the SF36 Health Survey. The authors showed that the beta-binomial regression approach renders satisfactory results in a broad number of situations, with a very convenient clinical interpretation of the results.

Finally, recoding of the scores to a binomial form is necessary in order to fit the beta-binomial

distribution to the SF-36 Health Survey scores. The same authors proposed and evaluated a method of recoding continuous and bounded scores, such as HRQoL scores, to a binomial form (Arostegui et al., 2013). The method was mainly based on the possible number of values each dimension can obtain, which, as it has been explained in Section 3.1 and showed in Table 4, comes from the number of items related with the construction of each dimension. Indeed, the methodology transforms the sorted scale of possible values of each dimension score to a ordinal scale from 0 to m, i.e. to a binomial form scale. The real interval [0,100], which is the scale of the standardized original scores, is divided in some subintervals, and then, each subinterval is linked to the value that corresponds in the order of the $0 - m$ scale, where $m + 1$ is the number of intervals. Consequently, score values within each subinterval are recoded with the value the subinterval was linked in the $0 - m$ scale. The way the subintervals are constructed is the main contribution of Arostegui et al. (2013)Arostegui et al. (2013) and the subdivision of the $0 - 100$ scale for each dimension is available in the Appendix of the mentioned work. We present in Table 5 the recoding process, step by step, for the dimensions that we will analyze in Section 3.6. Table 5 shows the possible values of the raw and standardized scores, as well as the subinterval division of the standardized original scores and the final recoded scores for *physical functioning*, *mental health* and *role emotional* dimensions. The authors evaluated and validated the proposed method of recoding with the scores provided by the SF-36 Health Survey. They showed that the recoding has a natural interpretation, not only for ordinal scores, but also for questionnaires with many dimensions and different profiles, where a common method of analysis is desired, such as the SF-36. Briefly, let Y denote the original standardized score observed in $[0, 100]$ and Y^* the recoded ordinal score, from 0 to m, where $Y^* \sim \text{Bin}(m, p)$. Thus, Y^* could be interpreted as grouped data for a dichotomous outcome that represents the number of successes in m binomial trials and p represents the probability of success in each trial. In the HRQoL context, Y^* is interpreted as the number of "points" that an individual has and p as the probability of obtaining one point more.

[Table 5 about here.]

Then, we will show the results obtained after applying the recoding to the SF-36 sores from the COPD Study data. Figure 1 shows the distribution of the eight recoded HRQoL dimensions in patients with COPD, together with the fit by the binomial and beta-binomial distributions. Similar shape can be observed for the original scores (Figure 8) and the recoded scores (Figure 1) although represented in different scales.

[Figure 1 about here.]

Figure 1 illustrates that the distributions of the recoded SF-36 dimensions are, generally, very skewed, accumulating values at the boundaries. As it can be seen the binomial distribution offers a poor fit in most recoded dimensions (e.g. role physical, body pain, social functioning, role emotional and mental health). Figure 1 also shows that the dimensions have different shapes (e.g.: bell-shaped, U or J-shaped), due to the fact that in some dimensions people tend to answer more or less extreme than in others. Consequently, there is an individual within variability in each dimension, that as it can be appreciated, the beta-binomial distribution is able to fit.

Figure 2 shows in a descriptive way the distribution of the eight HRQoL dimensions provided by the SF-36 of patients with COPD though different categorical variables, such as gender, dyspnea, anxiety and depression. It allows the descriptive analysis of the influence of each categorical characteristic in HRQoL of patients. Each axis of the radar chart corresponds to a recoded SF-36 dimension. The scales have been standardized to the interval defined by the length of the axis and divided into three cut points (25%, 50% and 75%) for a better visualization of the mean values. In Figure 2b we can appreciate the influence of the dyspnea in the eight HRQoL dimensions, where lower levels of dyspnea are associated to higher HRQoL in all the

dimensions. However, it can be shown that the mean effect of different dyspnea levels is not equal in all the dimensions, as the effects in *physical functioning* dimension is higher than in *body pain*, differentiating category effect between physical and mental dimensions. Figure 2a also shows that the mean perception of HRQoL is better in males than in females in all the dimension, being the mental dimensions where the difference was higher. On the other hand, Figures 2c-2d show that, as expected, an anxiety or depression status worsens in average the HRQoL of COPD patients in all the dimensions, the anxiety specially in the *role emotional* dimension and the depression in the *vitality*.

[Figure 2 about here.]

3.6   Modeling Results

Both approaches presented in Section 2, namely the *BBlogit* and the *hglm*, were applied to data from the COPD Study. The eight dimensions of the SF-36 Health Survey were the response variables and clinical and socio-demographic variables listed before were considered as independent variables. Independent models were performed for each of the health dimension of the SF-36 and exclusively data from the first visit to the outpatient clinic was considered. For variables selection, we retained in the model those covariates whose influence in HRQoL was statistically significant ($p < 0.05$) in at least one of the models. For simplicity, clarity and brevity of exposition, we only show results for three of the eight health dimensions of the SF-36. The selected three dimensions (*physical functioning*, *mental health* and *role emotional*) illustrate different shapes of the distribution and a wide range of maximum number of scores (m), from 4 to 20.

In terms of statistical packages and software we have implemented the beta-binomial logistic link approach in the R package HRQoL available at https://github.com/josunajera/HRQoL. This package unifies different analysis approaches for HRQoL data from SF-36 Health Survey and provides useful tools (such as specific quality of life radar charts). For the HGLM approach we used the R package hglm Rönnergard et al. (2010).

Results obtained from the beta-binomial regression analysis performed with the COPD study sample are provided in Tables 6-8. Estimates of the regression coefficients, their standard deviations and test of significance associated to the *BBlogit* and the *hglm* modelling approaches for the selected three health dimensions of the SF-36 Health Survey are displayed. We also show the estimates of the dispersion parameter of each model, α for *hglm* from equation (10), and ϕ for *BBlogit* from equation (1), in logarithmic scale and its significance test p-value.

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

Real data application leads to several conclusions and interpretations. As regards the interpretation of the fixed part regression coefficients β in both models is equivalent to the log odds-ratio in a binomial logistic regression model. For instance, the coefficient of depression in the *physical functioning* model for the *BBlogit* approach is -0.544, which means that based on this model the presence of depression increases by $1/exp(-0.544)=1.72$ the odds of having a smaller *physical functioning* score. In term of the interpretation based on the original SF-36 score, with a range from 0 to 100, it means that the presence of depression increases by 1.72 the odds of having 5 points less in the *physical functioning* domain. However, the most important issue we address in this work is that both methodologies lead to contradictory conclusions in terms of estimation and significance in some of the dimensions of the HRQoL provided by the SF-36. For instance, as shown in Table 6 in *physical functioning* dimension both methodologies

12

provided very similar results. However, in *mental health* (Table 7) and *role emotional* (Table 8) the results were different leading to different interpretations of the covariate effects. In *mental health* dimension the estimates of the coefficients were not very different, although they were not as similar as in *physical functioning* dimension, but interpretation problems occurred in terms of parameters significance. While the p-value corresponding to mild dyspnea in *hglm* was 0.134, leading that there was no difference between mild and no dyspnea, in *BBlogit* approach the effect was statistically significant reaching a p-value equal to 0.037. For *role emotional* dimension results were more contradictory, both the coefficient estimates and significances were very different. For example, on the one hand, the estimate of the coefficient corresponding to anxiety was -6.145 in *hglm* and -1.649 in *BBlogit*, being both statistically significant in the model. On the other hand, the p-value corresponding to the estimate of moderate dyspnea was statistically significant in *BBlogit* (<0.001), but not in *hglm* (0.434).

As shown in Section 2, due to the differences in constraint assumptions in both approaches and the non linear transformation of the mean response, it is expected to reach different regression parameter estimates. The *BBlogit* approach parameter estimation has a population-average interpretation while in *hglm* the interpretation is subject-specific, however the interpretation of the fixed part regression coefficients β in both models is equivalent to the log odds-ratio in a binomial logistic regression model. Consequently, the differences seem to be larger than we have previously expected comparing population-average and subject-specific models. Real data analysis also shows that those differences change through different shapes of the response variable. Moreover, Tables 6-8 show that standard errors of the estimates are also quite different leading to different parameter significance, and, consequently, contradictory results about the effect of the covariates in the HRQoL of the patients with COPD. In Table 8 it is shown that the anxiety is the unique statistically significant effect on the *role emotional* in *hglm* model, however, the estimated parameter variance is larger than 4, which makes us think that the estimates are over-inflated, leading to an incorrect model. Figure 3 shows the distribution of the modelled three HRQoL dimensions and the fit by the *hglm* model. It can be appreciated the subject-specific effect of the *hglm* model, in which the estimation is not focused on the mean and reaches almost all the score points in the analyzed three dimensions. Figure 3 also shows that although the estimates and variances of the parameters by *hglm* approach were confusing, the fit to the data is not so bad, specially in *role emotional* dimension, where the fit corresponds with the distribution of the dimension.

[Figure 3 about here.]

In general, it seems that when the value of the dispersion parameter increases, differences between both methodologies in terms of parameter estimation and significance become greater. The application of the two methodologies in real COPD data showed the need of the comparison of the two models that previously one might assume that will give more similar results, at least in terms of parameter significance. The need of a deeper analysis of the methodologies to understand why and when differences occur and to provide the best methodology in terms of covariate interpretation has been revealed by the real application. We believe that differences in the results depend on the dispersion parameter. Therefore, we focus on the comparison of the two approaches in the next section performing a complete simulation study dividing the analysis in scenarios depending on the value of the dispersion parameter.

4   Simulation study

4.1   Simulation scenarios

In this Section we set different scenarios in order to compare different methodologies to model HRQoL outcome data. Given the 8 health dimensions provided by the SF-36, we consider three

groups of possible maximum scores, m, i.e.: *few (4 and 3)*, *standard (8, 9 and 10)* and *large (19 and 20)* Arostegui et al. (2013) (as shown in Figure 1 for COPD data). Consequently, the simulation study has been also divided in three scenarios considering a maximum score of 4, 10 and 20. Finally, we have generated 500 random realizations of 100 observations of a dependent variable y assuming a beta-binomial distribution with fixed probability and dispersion parameters.

In order to understand the behaviour of the methodologies in the HRQoL framework, we are going to focus the simulation exercise on a regression approach with a single continuous covariate. The probability parameter has been calculated as shown in equation (6) for a fixed value of β_0 and β_1 equal to 1 and -0.3 respectively, and a fixed covariate simulated assuming a normal distribution with mean 3 and standard deviation 2.

The value of the dispersion parameter ϕ defines different scenarios (the shape of the beta distribution), as for a fixed probability parameter the shape of the distribution changes considerably for different values of ϕ. Values greater than 0.5 provides U-shaped distributions, values lower than 0.5 bell-shaped and a value equal to 0.5 flat-shaped. Figure 4 illustrates possible scenarios showing how the shape of the beta-binomial distribution changes for a fixed probability parameter equal to 0.5 considering the values $\phi = 2$ (Figure 4b), $\phi = 0.5$ (Figure 4c) and $\phi = 0.01$ (Figure 4d) for the dispersion parameter. If there is no over-dispersion, $\phi = 0$, it corresponds to the ordinary binomial distribution (Figure 4a). Hence, when the value of the dispersion parameter becomes greater the distribution is far from the mean value and the observations are accumulated at both extremes.

[Figure 4 about here.]

We implemented three methodologies in the simulation study (i) HGLM, (ii) beta-binomial regression with logistic link and (iii) mixed effects logistic regression model. The three methodologies have been deeply explained in Section 2.

4.2   Software implementation

The simulation process have been developed using different libraries in R (see Appendix B for details). The HGLM approach models have been performed using `hglm` library, which implements the estimation algorithm for hierarchical generalized linear models (Rönnergard et al., 2010). The rest of the models have been implemented using a library called `HRQoL` that has been developed by the authors. The function that performs the logistic regression based on a beta-binomial distribution is called `BBreg()` and the input is specified in a similar manner as for the standard `glm()` function in R. The remaining model in the simulation study, the GLMM, has been also performed with the so called `HRQoL` package by the `BIMreg()` function.

4.3   Results

Figures 5-7 show the box plots of the estimated regression parameters in the 500 simulations for the three methods of estimation: (i) HGLM (*hglm*), (ii) beta-binomial regression with logit link (*BBlogit*) and (iii) mixed-effects logistic regression (*glmm*). In order to compare the performance of the methods, Tables 9-11 include the mean, the variance and the expected mean square errors (EMS) of the estimated regression coefficients in each methodology. Tables 9-11 also show the percentage the real value of β_i is included in the 95% confidence interval of each simulation estimate (PCI) and the percentage the simulated covariate effect is statistically significant in each model (PCSS).

[Table 9 about here.]

[Table 10 about here.]

14

[Table 11 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

The simulation study shows that results of the three methodologies differ for different values of the dispersion parameter and different maximum scores as occurred in the COPD analysis. Generally, results provided by the *hglm* approach are more different to the real ones when the maximum number of scores and the dispersion parameter increase. In this situations, the regression parameter estimates have more bias than the other approaches due to the constraint assumption it assumes, which does not allow for any conclusions about the adequacy of the methodology. However, as the dispersion parameter increases variance of the estimates becomes larger, increasing the uncertainty of the parameter estimates. Moreover, due to the large variability of the estimates, the significance tests of covariate effect fail, reaching in same cases a PCSS of 22.6%, which is very low. Consequently, when the dispersion parameter is large enough, the *hglm* approach is not able to capture the effect of the covariate adequately. Consequently, patients' and disease's characteristics that influence the HRQoL dimensions are not going to be detected correctly by *hglm* approach.

On the other hand, the *BBlogit* approach gets more stable results in terms of the variance of the estimates and significance test than the *hglm* in all the simulated scenarios. Moreover the obtained estimates have less bias than both *hglm* and *glmm* methodologies. Furthermore, although when the dispersion parameter is large enough, in U-shaped scenario (Table 11), the variance of the estimates is not too large, but specially, this methodology continues capturing the effect of the covariate in the outcome variable correctly, reaching in the worst case a PCSS value equal to 71.4%.

Finally, the results provided by *glmm* approach also lead to several conclusions. Firstly, in the bell-shaped scenario (Table 9), as expected by the model assumption, the results provided by *glmm* approach, which unlike the other methodologies it does not depend on the beta-binomial distribution, get similar results to the *hglm* and *BBlogit*. However, as the dispersion parameter increases, model assumption begins to fail, as the random effects are not gaussian, and consequently, results get more unstable, increasing the variance of the estimates (Tables 10-11). Although there are enough arguments, as explained before, to conclude that *glmm* approach is not adequate to model HRQoL data, it should be emphasized that the PCSS is almost as good as in *BBlogit* approach. Consequently, although the estimates and variances of the regression parameter are not adequate, this methodology measures the statistically significant effect of the covariate correctly.

5   Discussion

We have illustrated that the use of two different approaches for the beta-binomial regression analysis may lead to different interpretations and statistical significances of the regression coefficients. On the one hand, apparently, results from *hglm* approach lead to an incorrect estimation of the β coefficients, however we must take into account the structure of the model. By defining random effects models we must impose constraints either on the fixed effects or in the random effects to maintain invariance between different modelling approaches. In *hglm* beta-binomial regression constraints are imposed in the random components as it was shown in equation (9), assuming that they have expectation equal to 1/2. The linear predictor shown in equation (8) is composed by the inclusion of random components with a logit transformation, which does not maintain the linearity property of the expectation operator, leading to a linear predictor with

some random effects that have no zero mean. The expectation of the random effects becomes greater as the dispersion parameter becomes larger, and as the linear predictor must be correctly estimated (as shown in Figure 3), differences between real and estimated β's increase. On the other hand, when random components are quite dispersed, a little variation in their distribution causes a big impact on the transformation through the logit link function, changing the mean of the random effects considerably. Hence, similar models can have different regression parameter estimates as we have shown in the simulation study, where the variance of the estimates in U-shaped scenario was very large. Furthermore, the impact that the prediction of the random effects has in the regression parameter estimates is quite big, which enlarges the uncertainty of regression parameters estimation, increasing the standard deviation of the estimates $\hat{\beta}$ and leading to erroneous significance tests. Consequently, although the *hglm* is an adequate model, as shown in Figure 3, to fit HRQoL, it is no valid to interpret covariate effects.

From a practical point of view, researchers working on HRQoL must be provided with a valid method of analysis for this kind of data. In some cases, the interest is focused on looking for an adequate fit to the HRQoL data, but other studies are focused on the effect of several covariates on HRQoL in a regression fashion. Our results showed that when the goal of the study is to detect and interpret the effect of explanatory variables in HRQoL the method of analysis must be cautiously selected. Therefore, we recommend the use of the *BBlogit* approach to perform beta-binomial regression of HRQoL data and covariate effect analysis on HRQoL data. Moreover, the `HRQoL` package in `R` provides the researchers the framework to perform *BBlogit* analysis for HRQoL data measured with the SF-36 Health Survey as a whole process, including the recoding of the original SF-36 scores, the descriptive analysis, the fit to the beta-binomial distribution and the beta-binomial regression.

Finally, we have evaluated two different approaches to perform beta-binomial regression of HRQoL as response variable and presented the results obtained for three specific dimensions of the SF-36 health questionnaire in stable COPD patients. Alternative dimensions of the SF-36, different HRQoL questionnaires or other populations may have different characteristics, and therefore, methods may perform in a different way. However, many other HRQoL data and, in general, patient-reported outcomes share many characteristics that make them suitable for the methodological approaches presented here.

Acknowledgements

Appendix A

We include in this Appendix some descriptive statistics of the original standardized scores of the SF-36 in order to show the original data and its distribution. We considered Figure 8 and Table 12 as supplementary material, which are not necessary for the correct understanding of the applied methodology, but still helps the reader for an easier comprehension of the modeling proposal.

[Figure 8 about here.]

[Table 12 about here.]

Appendix B

We include a short code of the simulation study of the flat-shaped scenario for a maximum number of score equal to 4 in `R`.

```
# Load the required R-packages
require(hglm)
require(HRQoL)


beta <- c(1,-0.3) # Define the regression coefficients
phi <- 0.5 # Define the dispersion parameter (Flat-shaped)

nSim <- 500 # Number of simulations
nObs <- 100 # Number of observations in each simulation
nTrial <- 4

set.seed(1)
x <- rnorm(nObs,3,2) # The covariate
rm(list=".Random.seed", envir=globalenv()) # Remove the set.seed()

X <- cbind(1,x) # The design matrix of the regression coefficients
id <- seq(1,nObs,1) # The random components

p <- 1/(1+exp(-(X%*%beta)))  # The probability parameter for the given
    parameters and covariate

 for(i in 1:nSim){
    y <- rBB(nObs,nTrial,p,phi) # Simulate the response variable
    dat <- data.frame(cbind(y,x,id)) # Create the data frame
    dat$id <- as.factor(dat$id) # Random components as factors

    # The 3 approaches
    hglm <- hglm(fixed=y/nTrial~x, random = ~1|id, weights=rep(nTrial,nObs
        ), family = binomial(link=logit), rand.family = Beta(link=logit),
        fix.disp = 1) # The hglm model

    BBLogit <- BBreg(y~x,n=nTrial,data=dat) # The BBlogit model

    glmm <- BIMreg(y~x,random=c("id"),n=nTrial,data=dat) # The glmm model

        # The regression coefficients
        coef.hglm <- cbind(coef.hglm,hglm$fixef)
        coef.BBlogit <- cbind(coef.BBlogit,BBlogit$coef)
        coef.glmm <- cbind(coef.glm,glmm$coefficients)

        # The standard errors of the estimations
        se.hglm <- cbind(se.hglm,hglm$SeFe)
        se.BBlogit <- cbind(se.BBlogit,sqrt(diag(BBlogit$vcov)))
        se.glmm <- cbind(se.glmm,sqrt(diag(glmm$coef.vcov)))

        # The variance parameters of the random effects
        hglm.lambda <- c(hglm.lambda,hglm$varRanef)
        BBlogit.phi <- c(BBlogit.phi, BBlogit$phi.coefficient)
        glmm.sigma <- c(glmm.sigma,glmm$random.se)

        # The random effects
```

```
        hglm.rand <- cbind(hglm.rand,hglm$ranef)
        glmm.rand <- cbind(glmm.rand,glmm$random.coef)
}
```
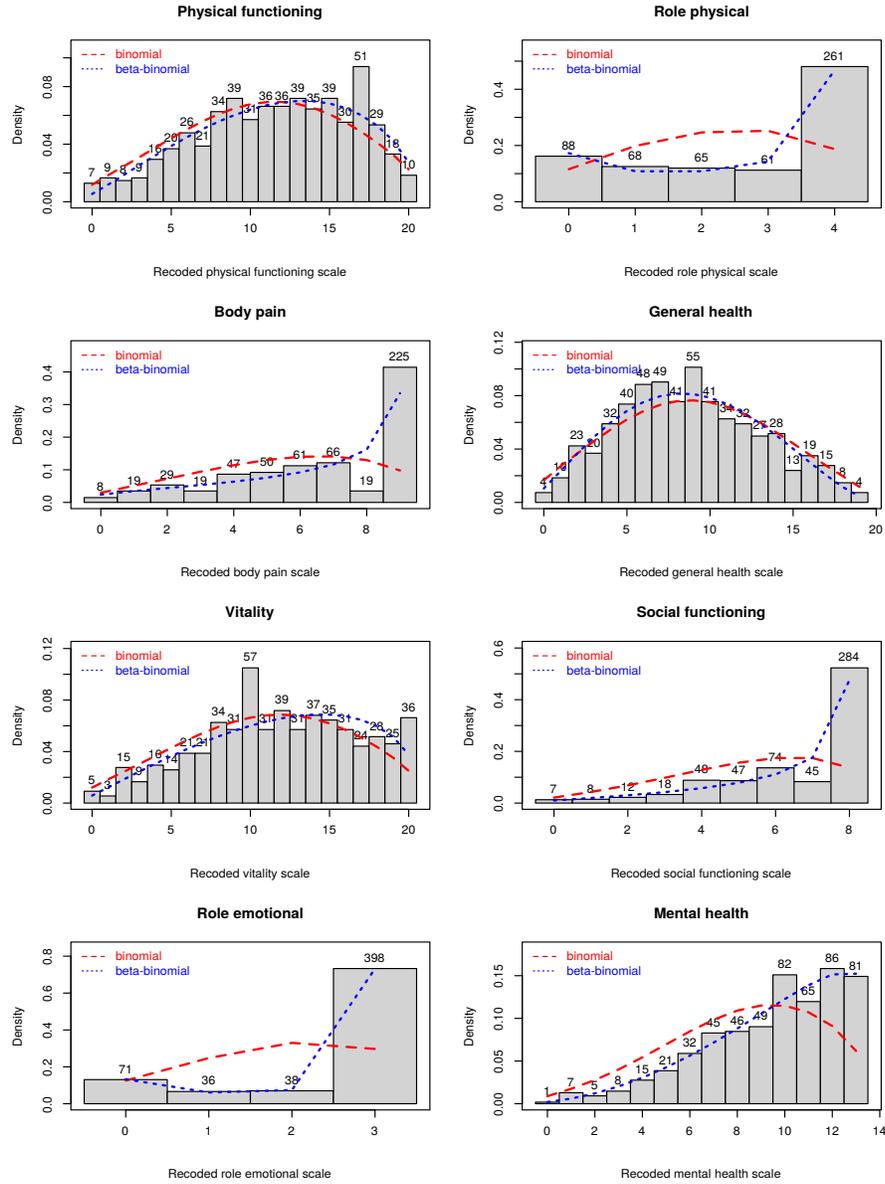
References

Testa MA, Simonson DC. Assessment of quality of life outcomes. *The New England Journal of Medicine* 1996; **334** : 835-840.

Goldsmith SB. The status of health status indicators. *Health Services Reports* 1972; **87** : 212-220.

Ware JE, Snow KK, Kosinski MA, Gandek B. *SF36 Health Survey, Manual and Interpretation Guides.* The Health Institute, New England Medical Center, 1993.

Arostegui I, Nuñez-Antón V, Quintana JM. Analysis of the Short Form-36 (SF-36): The beta-binomial distribution approach. *Statistics in Medicine* 2007; **26** : 1318-1342.

Arostegui I, Nuñez-Antón V, Quintana JM. Statistical approach to analyse patient-reported outcomes as response variables: An application to health-related quality of life. *Statistical Methods in Medical Research* 2012; **21** : 189-214.

Forcina A, Franconi L. Regression analysis with the beta-binomial distribution. *Rivista di Statistica Applicata* 1988; **21** : 7-12.

Lee Y, Nelder JA. Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B* 1996; **58** : 619-678.

Lee Y, Nelder JA, Noh M. H-likelihood: problems and solutions. *Statistics and Computing* 2007; **17** : 49-55.

Lee W, Lee Y. Modifications of REML algorithm for HGLMs. *Statistics and Computing* 2012; **22** : 959-966.

McCulloch CE, Searle SR. *Generalized, Linear and Mixed Models.* John Wiley & Sons, INC., 2001.

McCullagh P, Nelder JA. *Generalized Linear Model.* Chapman & Hall, 1989.

Lee Y, Nelder JA, Pawitan Y. *Generalized Linear Models with Random Effects: Unified Approach via h-Likelihood.* Chapman & Hall, 2006.

Lee Y, Nelder JA. Hierarchical generalized linear models: A synthesis of generalized linear models, random-effects and structured dispersions. *Biometrika* 2001; **88** : 987-1006.

Schall R. Estimation in generalized linear models with random effects. *Biometrika* 1991; **78** : 719-727.

Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993; **88** : 9-25.

Halbert RJ, Isonaka S, George D, Iqbal A. Interpreting COPD prevalence estimates: What is the true burden of disease? *Chest* 2003; **123** : 1684-92.

Buist AS, Vollmer WM, McBurnie MA. Worldwide burden of COPD in high-and low-income countries. Part I. The burden of obstructive lung disease (BOLD) initiative. *The International Journal of Tuberculosis and Lung Disease* 2008; **12** : 703-708.

Murray CJ, López AD. Alternative projections of mortality and disability by cause 1990-2020: Global burden of disease study. *Lancet* 1997; **349** : 1498-504.

Esteban C, Quintana JM, Moraza J, Aburto M, Egurrola M, España PP, Pérez-Izquierdo J, Aguirre U, Aizpiri S, Capelastegui A. Impact of hospitalisations for exacerbations of COPD on health-related quality of life. *Respiratory Medicine* 2009; **103** : 1201-1208.

Mahler D, Ward J, Waterman L, McCusker C, Zuwallack R, Baird J. Patient-reported dyspnea in COPD reliability and association with stage of disease. *Chest* 2009; **136** : 1473-1479.

ATS statement: guidelines for the six-minute walk test. *American Journal of Respiratory and Critical Care Medicine* 2002; **166** : 11-117.

Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica* 1983; **67** : 361-370.

Stansfeld SA, Roberts R, Foot SP. Assessing the validity of the SF-36 general health survey. *Quality of Life Research* 1997; **6** : 217-224.

SF-36 Health Survey. *Manual and interpretation guide.* Boston, MA: The Health Institute, New England Medical Center, 1993.

Ware JE, Kosinski MA, Keller SD. *SF-36 Physical and Mental Health Summary Scales: A Users Manual.* Boston: The Health Institute, New England Medical Center, 1994.

Garratt AM, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: Bibliographic study of patient assessed health outcome measures. *British Medical Journal* 2002; **324** : 1417-1421.

Arostegui I, Nuñez-Antón V, Quintana JM. On the recoding of continuous and bounded indexes to a binomial form: an applications to quality-of-life scores. *Journal of Applied Statistics* 2013; **40** : 563-582.

Rönnergard L, Shen X, Alam M. hglm: A package for fitting hierarchical generalized linear models. *The R Journal* 2010; **2** : 20-28.
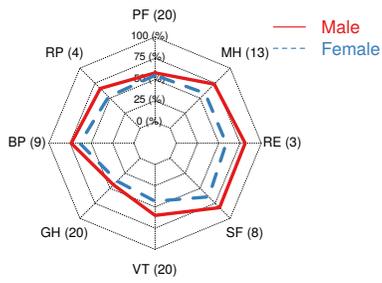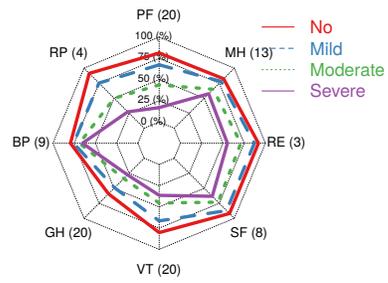
List of Figures

Figure 1: Histogram and fit with binomial and beta-binomial distributions of the recoded HRQoL dimensions in COPD data. Frequencies are shown at the top of each bar.

(a) Distribution of the HRQoL by the gender

(b) Distribution of the HRQoL by the dyspnea status
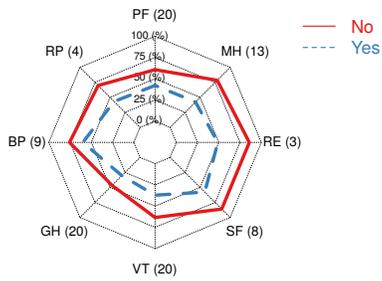
(c) Distribution of the HRQoL by the anxiety
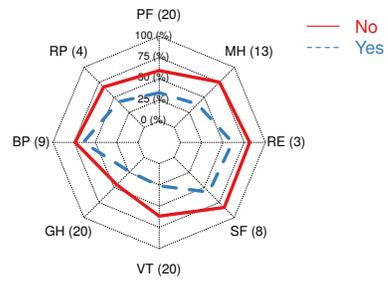
(d) Distribution of the HRQoL by the depression

Figure 2: Distribution of HRQoL in COPD patients through some categorical variables. Numbers between parenthesis are the maximum values of the recoded SF-36 scores in each dimension.
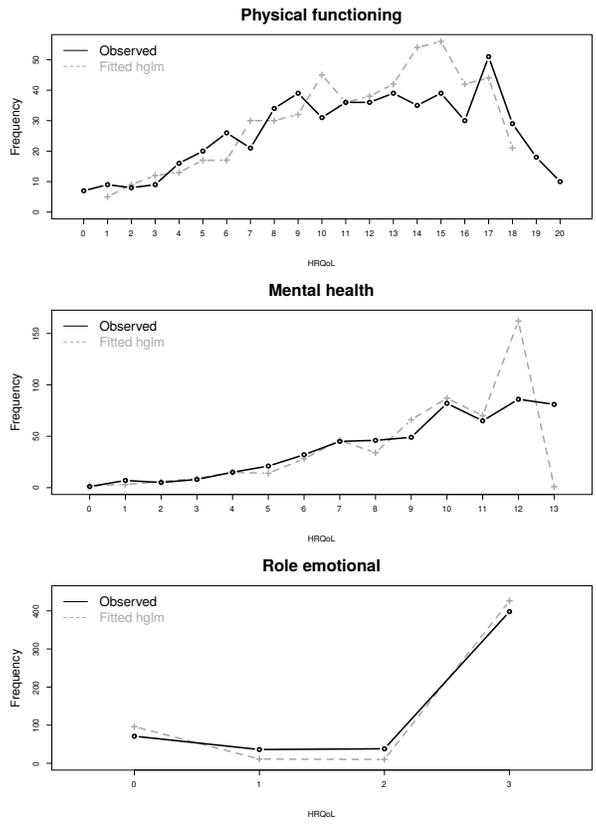
Figure 3: Observed distribution and fitted distribution by the *hglm* approach of HRQoL scores.
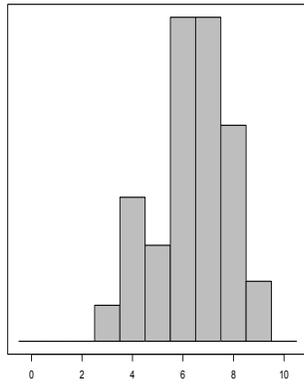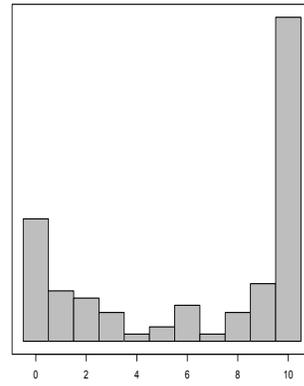
(a) Binomial ($\phi = 0$)   (b) U-shaped ($\phi = 2$)
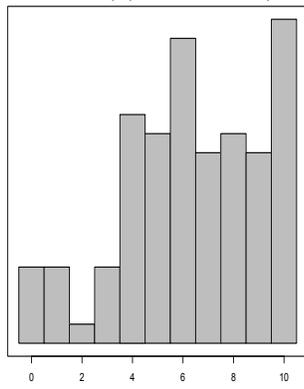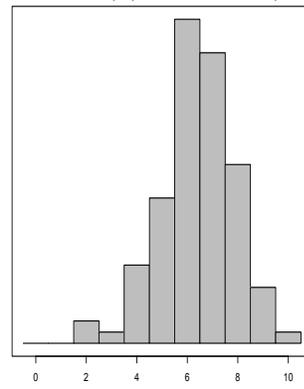
(c) Flat-shaped ($\phi = 0.5$)   (d) Bell-shaped ($\phi = 0.01$)

Figure 4: Different scenarios based on parameter ϕ.
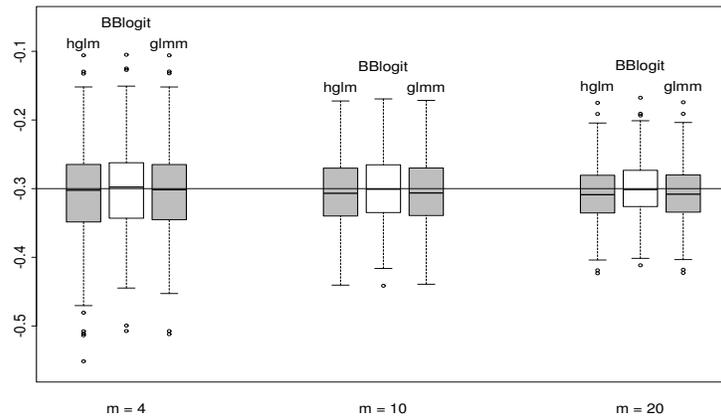
Figure 5: Box plots of the slope estimates in the simulation study for bell-shaped scenario ($\phi = 0.01$). Simulations performed for $n = 100$ individuals and $R = 500$ replicates.
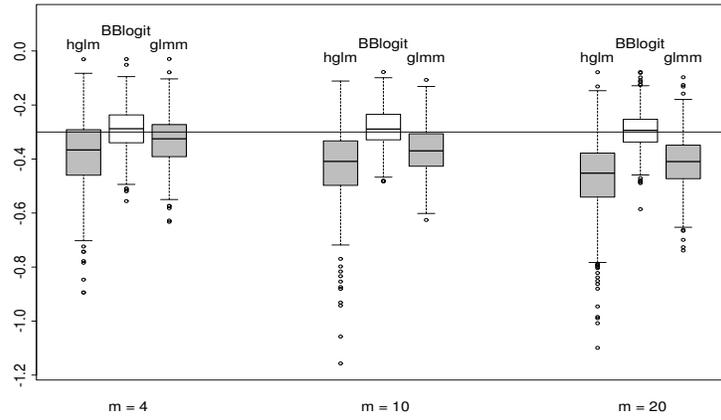
Figure 6: Box plots of the slope estimates in the simulation study for flat-shaped scenario ($\phi = 0.5$). Simulations performed for $n = 100$ individuals and $R = 500$ replicates.
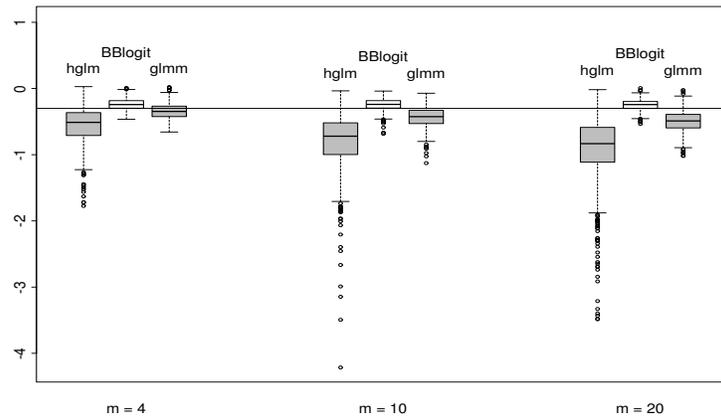
Figure 7: Box plots of the slope estimates in the simulation study for U-shaped scenario ($\phi = 2$). Simulations performed for $n = 100$ individuals and $R = 500$ replicates.
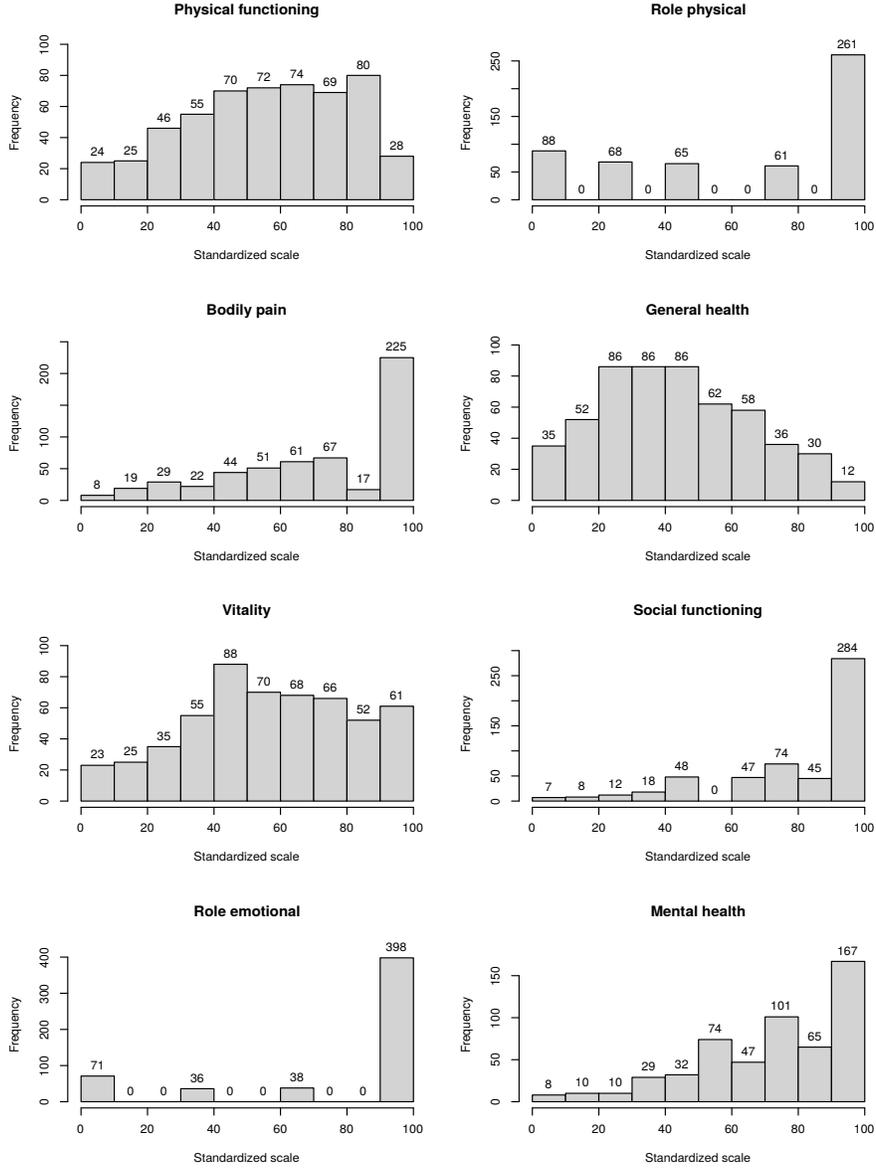
Figure 8: Distribution of the original standardized scores of the eight SF-36 dimensions in COPD data. Frequencies are shown at the top of each bar.

List of Tables

Table 1: Descriptive analysis of the socio-demographic and clinical variables. The total number of individuals is $n = 543$.

Discrete variables	n (%)
Dyspnea	
No	69 (12.71%)
Mild	264 (48.62%)
Moderate	166 (30.57%)
Severe	44 (8.10%)
Depression	
No	506 (93.18%)
Yes	37 (6.82%)
Anxiety	
No	459 (84.53%)
Yes	84 (15.47%)
Continuous variables	**Mean (SD)**
FEV1%	55 (13.31)
BMI	28.27 (4.43)
Age	68.32 (8.32)
Walking Test	408.89 (92.43)

SD: Standard Deviation; BMI: Body Mass Index; FEV1%: Forced Expiratory Volume in one second in percentile.

Table 2: Descriptive analysis of the original standardized scores of the eight SF-36 dimensions in COPD data.

Dimension	Mean	SD	No. of items	No. of possible values
Physical functioning	57.762	24.382	10	21
Role physical	65.608	38.916	4	5
Body pain	71.087	29.259	2	27
General health	44.672	21.934	5	39
Vitality	59.355	24.957	4	21
Social functioning	81.534	24.460	2	9
Role emotional	80.172	35.905	3	4
Mental health	73.422	22.918	5	26

SD: Standard Deviation.

Table 3: Recoding methodology for *role emotional*, *mental health* and *physical functioning* dimensions.

Role emotional

Raw	Stdr.	Inter.	Recoded
3	0	[0, 16.67]	0
4	33.3	(16.67, 50]	1
5	66.7	(50, 83.33]	2
6	100	(83.33, 100]	3

Physical functioning

Raw	Stadr.	Inter.	Recoded
10	0	[0, 2.5]	0
11	5	(2.5, 7.5]	1
12	10	(7.5, 12.5]	2
13	15	(12.5, 17.5]	3
14	20	(17.5, 22.5]	4
15	25	(22.5, 27.5]	5
16	30	(27.5, 32.5]	6
17	35	(32.5, 37.5]	7
18	40	(37.5, 42.5]	8
19	45	(42.5, 47.5]	9
20	50	(47.5, 52.5]	10
21	55	(52.5, 57.5]	11
22	60	(57.5, 62.5]	12
23	65	(62.5, 67.5]	13
24	70	(67.5, 72.5]	14
25	75	(72.5, 77.5]	15
26	80	(77.5, 82.5]	16
27	85	(82.5, 87.5]	17
28	90	(87.5, 92.5]	18
29	95	(92.5, 97.5]	19
30	100	(97.5, 100]	20

Mental health

Raw	Stdr.	Inter.	Recoded
5	0	[0,2]	0
6	4	(2,10]	1
7	8		
8	12	(10,18]	2
9	16		
10	20	(18, 26]	3
11	24		
12	28	(26,34]	4
13	32		
14	36	(34,42]	5
15	40		
16	44	(42,50]	6
17	48		
18	52	(50,58]	7
19	56		
20	60	(58,66]	8
21	64		
22	68	(66,74]	9
23	72		
24	76	(74,82]	10
25	80		
26	84	(82,90]	11
27	88		
28	92	(90,98]	12
29	96		
30	100	(98,100]	13

Raw: Raw scores; Stdr.: Standardized original scores; Inter.: The subinterval division of the $0-100$ scale; Recoded: Recoding of the values. The decomposition in raw and original standardized scores of the HRQoL dimensions is developed in Ware et al. (1993) (Ware et al., 1993), while the subinterval division and recoding process is explicitly explained in Arostegui et al. (2013) (Arostegui et al., 2013).

Table 4: Descriptive analysis of the original standardized scores of the eight SF-36 dimensions in COPD data.

Dimension	Mean	SD	No. of items	No. of possible values
Physical functioning	57.762	24.382	10	21
Role physical	65.608	38.916	4	5
Body pain	71.087	29.259	2	27
General health	44.672	21.934	5	39
Vitality	59.355	24.957	4	21
Social functioning	81.534	24.460	2	9
Role emotional	80.172	35.905	3	4
Mental health	73.422	22.918	5	26

SD: Standard Deviation.

Table 5: Recoding methodology for *role emotional*, *mental health* and *physical functioning* dimensions.

Role emotional

Raw	Stdr.	Inter.	Recoded
3	0	[0, 16.67]	0
4	33.3	(16.67, 50]	1
5	66.7	(50, 83.33]	2
6	100	(83.33, 100]	3

Physical functioning

Raw	Stadr.	Inter.	Recoded
10	0	[0, 2.5]	0
11	5	(2.5, 7.5]	1
12	10	(7.5, 12.5]	2
13	15	(12.5, 17.5]	3
14	20	(17.5, 22.5]	4
15	25	(22.5, 27.5]	5
16	30	(27.5, 32.5]	6
17	35	(32.5, 37.5]	7
18	40	(37.5, 42.5]	8
19	45	(42.5, 47.5]	9
20	50	(47.5, 52.5]	10
21	55	(52.5, 57.5]	11
22	60	(57.5, 62.5]	12
23	65	(62.5, 67.5]	13
24	70	(67.5, 72.5]	14
25	75	(72.5, 77.5]	15
26	80	(77.5, 82.5]	16
27	85	(82.5, 87.5]	17
28	90	(87.5, 92.5]	18
29	95	(92.5, 97.5]	19
30	100	(97.5, 100]	20

Mental health

Raw	Stdr.	Inter.	Recoded
5	0	[0,2]	0
6	4	(2,10]	1
7	8		
8	12	(10,18]	2
9	16		
10	20	(18, 26]	3
11	24		
12	28	(26,34]	4
13	32		
14	36	(34,42]	5
15	40		
16	44	(42,50]	6
17	48		
18	52	(50,58]	7
19	56		
20	60	(58,66]	8
21	64		
22	68	(66,74]	9
23	72		
24	76	(74,82]	10
25	80		
26	84	(82,90]	11
27	88		
28	92	(90,98]	12
29	96		
30	100	(98,100]	13

Raw: Raw scores; Stdr.: Standardized original scores; Inter.: The subinterval division of the $0-100$ scale; Recoded: Recoding of the values. The decomposition in raw and original standardized scores of the HRQoL dimensions is developed in Ware et al. (1993) (Ware et al., 1993), while the subinterval division and recoding process is explicitly explained in Arostegui et al. (2013) (Arostegui et al., 2013).

Table 6: Effect of explanatory variables in *physical functioning* dimension measured by beta-binomial regression and estimated by both approaches.

Physical functioning	hglm			BBlogit		
	$\hat{\beta}$	SD($\hat{\beta}$)	p-value	$\hat{\beta}$	SD($\hat{\beta}$)	p-value
Dyspnea						
Mild	-0.616	0.111	<0.001	-0.580	0.112	<0.001
Moderate	-1.339	0.122	<0.001	-1.281	0.120	<0.001
Severe	-2.317	0.178	<0.001	-2.207	0.176	<0.001
Depression						
Yes	-0.541	0.139	<0.001	-0.544	0.130	<0.001
Anxiety						
Yes	-0.416	0.096	<0.001	-0.404	0.090	<0.001
Sex						
Female	0.469	0.167	0.005	0.461	0.155	0.003
FEV1%	0.007	0.003	0.011	0.006	0.002	0.012
BMI	-0.019	0.007	0.011	-0.018	0.007	0.009
Age	0.013	0.004	0.002	0.012	0.004	0.002
Walking Test	0.004	10^{-4}	<0.001	0.004	10^{-4}	<0.001
$\log(\alpha)$	-2.656	0.084	<0.001	—	—	—
$\log(\phi)$	—	—	—	-2.826	0.115	<0.001

SD: Standard Deviation; BMI: Body Mass Index; FEV1%: Forced Expiratory Volume in one second in percentile.

Table 7: Effect of explanatory variables in *mental health* dimension measured by beta-binomial regression and estimated by both approaches.

Mental health	hglm			BBlogit		
	$\hat{\beta}$	SD($\hat{\beta}$)	p-value	$\hat{\beta}$	SD($\hat{\beta}$)	p-value
Dyspnea						
Mild	-0.353	0.234	0.134*	-0.294	0.141	0.037
Moderate	-0.853	0.246	<0.001	-0.704	0.145	<0.001
Severe	-1.132	0.320	<0.001	-0.961	0.181	<0.001
Anxiety						
Yes	-1.480	0.204	<0.001	-1.290	0.108	<0.001
Depression						
Yes	-0.966	0.298	0.002	-0.853	0.157	<0.001
$\log(\alpha)$	-0.7647	0.069	<0.001	—	—	—
$\log(\phi)$	—	—	—	-2.263	0.115	<0.001

SD: Standard Deviation. Symbol * stands for regression coefficients that were not statistically significant.

Table 8: Effect of explanatory variables in *role emotional* dimension measured by beta-binomial regression and estimated by both approaches.

Role emotional	hglm			BBlogit		
	$\hat{\beta}$	SD($\hat{\beta}$)	p-value	$\hat{\beta}$	SD($\hat{\beta}$)	p-value
Anxiety						
Yes	-6.145	2.062	0.003	-1.649	0.226	<0.001
Dyspnea						
Mild	-2.600	5.229	0.619*	-0.614	0.418	0.142*
Moderate	-3.981	5.080	0.434*	-1.379	0.413	<0.001
Severe	-5.603	5.496	0.309*	-2.048	0.467	<0.001
$\log(\alpha)$	2.735	0.095	<0.001	−	−	−
$\log(\phi)$	−	−	−	0.668	0.150	<0.001

SD: Standard Deviation. Symbol * stands for regression coefficients that were not statistically significant.

Table 9: Results of the simulation study for the bell-shaped distribution ($\phi = 0.01$) for $n = 100$ individuals and $R = 500$ replicates.

True value	Method	$\beta_0 = 1$			$\beta_1 = -0.3$			
		Mean(SD)	EMS	PCI	Mean(SD)	EMS	PCI	PCSS
$m = 4$	hglm	1.024 (0.245)	0.060	95.8%	-0.309 (0.067)	0.004	97.4%	100%
	BBlogit	1.007 (0.228)	0.052	95.6%	-0.304 (0.063)	0.004	97.2%	100%
	glmm	1.015 (0.230)	0.053	95.8%	-0.306 (0.063)	0.004	97.6%	100%
$m = 10$	hglm	1.018 (0.173)	0.030	96.0%	-0.305 (0.048)	0.002	95.8%	100%
	BBlogit	1.001 (0.169)	0.028	95.6%	-0.299 (0.047)	0.002	96.0%	100%
	glmm	1.016 (0.170)	0.029	95.6%	-0.304 (0.047)	0.002	95.8%	100%
$m = 20$	hglm	1.030 (0.140)	0.020	94.8%	-0.308 (0.038)	0.001	95.2%	100%
	BBlogit	1.004 (0.136)	0.019	95.4%	-0.304 (0.037)	0.001	96.0%	100%
	glmm	1.027 (0.140)	0.020	94.8%	-0.308 (0.038)	0.001	95.0%	100%

SD: Standard Deviation; EMS: Expected Mean Square errors; PCI: Percentage the real value is included in the 95% Confidence Interval of each simulation estimate; PCSS: Percentage the covariate effect is statistically significant.

Table 10: Results of the simulation study for the flat-shaped distribution ($\phi = 0.5$) for $n = 100$ individuals and $R = 500$ replicates.

True value		$\beta_0 = 1$			$\beta_1 = -0.3$			
	Method	Mean(SD)	EMS	PCI	Mean(SD)	EMS	PCI	PCSS
$m = 4$	hglm	1.284 (0.476)	0.307	96.4%	-0.384 (0.132)	0.024	95.6%	91.8%
	BBlogit	0.960 (0.280)	0.080	97.0%	-0.289 (0.078)	0.006	96.0%	92.6%
	glmm	1.100 (0.321)	0.113	96.6%	-0.330 (0.088)	0.009	96.4%	93.4%
$m = 10$	hglm	1.420 (0.514)	0.440	91.2%	-0.425 (0.138)	0.034	89.0%	96.0%
	BBlogit	0.953 (0.258)	0.069	95.8%	-0.287 (0.071)	0.005	95.0%	98.8%
	glmm	1.230 (0.331)	0.162	91.2%	-0.331 (0.091)	0.013	88.4%	98.6%
$m = 20$	hglm	1.574 (0.528)	0.610	86.8%	-0.468 (0.139)	0.048	83.2%	94.0%
	BBlogit	0.985 (0.252)	0.064	95.8%	-0.296 (0.064)	0.005	96.6%	99.2%
	glmm	1.370 (0.350)	0.259	81.0%	-0.411 (0.094)	0.021	80.0%	99.2%

SD: Standard Deviation; EMS: Expected Mean Square errors; PCI: Percentage the real value is included in the 95% Confidence Interval of each simulation estimate; PCSS: Percentage the covariate effect is statistically significant.

Table 11: Results of the simulation study for the U-shaped distribution ($\phi = 2$) for $n = 100$ individuals and $R = 500$ replicates.

True value		$\beta_0 = 1$			$\beta_1 = -0.3$			
	Method	Mean(SD)	EMS	PCI	Mean(SD)	EMS	PCI	PCSS
$m = 4$	hglm	1.809 (0.988)	1.631	97.4%	-0.559 (0.287)	0.150	92.6%	62.8%
	BBlogit	0.776 (0.294)	0.137	94.4%	-0.244 (0.086)	0.010	94.0%	71.4%
	glmm	1.105 (0.412)	0.181	99.0%	-0.347 (0.120)	0.016	97.6%	72.2%
$m = 10$	hglm	2.630 (1.648)	5.394	92.2%	-0.801 (0.467)	0.470	85.4%	39.6%
	BBlogit	0.777 (0.307)	0.144	92.8%	-0.244 (0.091)	0.011	91.2%	76.2%
	glmm	1.375 (0.528)	0.420	96.2%	-0.431 (0.156)	0.041	91.2%	75.2%
$m = 20$	hglm	3.210 (2.250)	9.946	86.2%	-0.951 (0.570)	0.749	81.6%	22.6%
	BBlogit	0.801 (0.313)	0.137	91.6%	-0.249 (0.086)	0.010	93.0%	82.0%
	glmm	1.594 (0.614)	0.730	88.4%	-0.498 (0.170)	0.068	84.4%	80.8%

SD: Standard Deviation; EMS: Expected Mean Square errors; PCI: Percentage the real value is included in the 95% Confidence Interval of each simulation estimate; PCSS: Percentage the covariate effect is statistically significant.

Table 12: Mean and standard deviation, in parenthesis, of the original standardized scores of the eight SF-36 dimensions stratified by categorical covariates of interest.

Covariates	n	PF	RP	BP	GH	VT	SF	RE	MH
Sex:									
Male	522	57.87	66.14	71.49	44.92	60.00	82.28	81.03	74.04
		(24.58)	(38.65)	(29.10)	(22.01)	(24.71)	(23.61)	(35.04)	(22.42)
Female	21	55.00	52.38	61.00	38.48	43.33	64.29	58.78	58.09
		(18.97)	(43.95)	(32.01)	(19.32)	(26.33)	(36.93)	(49.33)	(29.71)
Dyspnea									
No	69	81.23	91.30	78.09	60.52	80.21	92.03	91.79	83.30
		(18.08)	(22.62)	(30.17)	(19.41)	(19.40)	(18.56)	(24.52)	(18.86)
Mild	264	67.29	75.19	73.45	49.95	66.46	87.22	86.99	78.70
		(17.57)	(35.15)	(27.79)	(20.92)	(21.33)	(19.37)	(30.54)	(19.95)
Moderate	166	43.67	49.65	66.57	34.17	45.48	73.12	71.08	65.16
		(16.54)	(37.99)	(29.41)	(18.04)	(21.92)	(26.64)	(40.12)	(24.42)
Severe	44	16.93	27.27	62.93	27.79	36.36	63.35	55.30	57.45
		(11.11)	(33.16)	(32.36)	(15.88)	(19.95)	(31.48)	(44.28)	(21.94)
Depression:									
No	506	59.53	67.29	71.78	46.18	61.79	83.32	81.55	76.04
		(23.69)	(38.40)	(29.11)	(21.62)	(23.61)	(22.83)	(34.82)	(20.39)
Yes	37	33.51	42.57	61.68	23.93	26.08	57.77	61.26	37.62
		(20.68)	(39.03)	(29.97)	(14.60)	(18.26)	(32.72)	(44.80)	(25.78)
Anxiety:									
No	459	60.63	69.72	73.54	47.24	63.42	85.95	85.91	79.15
		(23.66)	(37.24)	(28.26)	(21.64)	(23.43)	(21.01)	(30.99)	(18.04)
Yes	84	42.08	43.15	57.68	30.61	37.14	57.74	48.81	42.09
		(22.32)	(40.39)	(30.59)	(17.92)	(21.13)	(28.11)	(44.04)	(21.44)

n: Number of individuals, PF: Physical functioning, RP: Role physical, BP: Body pain, GH: General health, VT: Vitality, SF: Social functioning, RE: Role emotional, MH: Mental health.