STATISTICS IN MEDICINE
Statist. Med. 2007; **26**:1318–1342
Published online 23 June 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2612

Analysis of the short form-36 (SF-36): The beta-binomial distribution approach

Inmaculada Arostegui[1,*,†], Vicente Núñez-Antón[2,‡] and José M. Quintana[3,§]

[1]*Departamento de Matemática Aplicada y Estadística e Investigación Operativa, Universidad del País Vasco (UPV/EHU), Bilbao, Spain*
[2]*Departamento de Econometría y Estadística (E.A. III), Universidad del País Vasco (UPV/EHU), Bilbao, Spain*
[3]*Unidad de Investigación Clínica, Hospital Galdakao, Bilbao, Spain*

SUMMARY

Health-related quality of life (HRQoL) is an important indicator of health status and the Short Form-36 (SF-36) is a generic instrument to measure it. Multiple linear regression (MLR) is often used to study the relationship of HRQoL with patients' characteristics, though HRQoL outcomes tend to be not normally distributed, skewed and bounded (e.g. between 0 and 100). A sample of 193 patients with eating disorders has been analysed to assess the performance of the MLR under non-normality conditions. Normal distribution was rejected for seven out of the eight domains. A beta-binomial distribution is suggested to fit the SF-36 scores. The beta-binomial distribution is not rejected for five out of the eight domains. Thus, a beta-binomial regression (BBR) is suggested to analyse the SF-36 scores. Results using MLR and BBR have been compared for real and simulated data. Performance of the BBR is shown to be better than MLR in the HRQoL domains with few ordered categories and very similar to MLR in the more continuous domains. Moreover, the interpretation of the estimates obtained with BBR is clinically more meaningful. A common technique of statistical analysis is preferable for all the HRQoL dimensions. Therefore, the BBR approach is recommended not only to detect significant predictors of HRQoL when SF-36 is used, but also to analyse and interpret the effect of several explanatory variables on HRQoL. Further work is required to test the better performance of BBR against standard methods for other HRQoL outcomes, populations or interventions. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: beta-binomial model; goodness-of-fit; health-related quality of life; SF-36

*Correspondence to: Inmaculada Arostegui, Departamento de Matemática Aplicada y Estadística e Investigación Operativa, Universidad del País Vasco (UPV/EHU), Aptdo. 644, E-48080 Bilbao, Spain.
†E-mail: inmaculada.arostegui@ehu.es
‡E-mail: vicente.nunezanton@ehu.es
§E-mail: jmquinta@hgda.osakidetza.net

Contract/grant sponsor: Universidad del País Vasco (UPV/EHU); contract/grant numbers: 1/UPV/EHU 00100.310-EA-8108/2000, UPV 00038.321-13631/2001
Contract/grant sponsor: Departamento de Sanidad del Gobierno Vasco and Ministerio Español de Educación y Ciencia and FEDER; contract/grant number: MTM2004-00341

Received 12 July 2005
Accepted 28 April 2006

1. INTRODUCTION

There has been an increasing recognition that typical clinical and laboratory outcomes are not necessarily the most important results in studies that examine the effect of health interventions. Therefore, great relevance has been placed on the importance of quality of life in measuring health status and evaluating medical care results, especially in chronic illnesses [1]. Measurement of the health-related quality of life (HRQoL) makes it possible to obtain information about illness and its impact on the patients' life, in a standardized, comparative and more objective way [2, 3]. Interest in HRQoL instruments started to increase during the decade of the eighties in four broad contexts: Measuring the health of populations, assessing the benefit of alternative uses of resources, comparing two or more interventions in a clinical trial and making a decision on treatment for an individual patient [4, 5]. Several HRQoL instruments have been developed in a questionnaire form, some of them are generic and others are disease specific. One of the most widely used is the Short Form-36 (SF-36), a generic HRQoL instrument consisting of 36 questions that has been designed and validated in healthy populations by Ware *et al.* [6]. In addition, it has been translated into several languages, including Spanish, and validated therein [7].

In the last decade, a lot of research has focused on HRQoL assessment in clinical trials [8, 9]. There is also work concentrated on the statistical analysis of HRQoL data in the context of clinical trials since the nineties [10–13]. In addition, the combined information about HRQoL and length of life in what is called quality-adjusted life years (QALYs) has been broadly used for health policy decisions and resource allocation [14]. However, less scientific research has been published in the other two contexts mentioned above, namely, measuring HRQoL of populations and making a decision on treatments for an individual patient. There are many settings where it is not possible to design a clinical trial and, thus, decisions on treatments are sometimes based on observational studies where HRQoL has been measured as outcome. The main goals of these studies involving HRQoL outcomes are not only to measure and compare HRQoL among different populations, but also to study the relationship of HRQoL with patients' and disease's characteristics. In many studies these are variables that are not controlled by the researcher during the study design, as they are controlled in clinical trials. We concentrate primarily on this context.

Multiple linear regression (MLR) is a frequently used multivariate technique for analysing the effect of several explanatory variables on HRQoL. Multiple regression models assume the following: (1) The relationship between the outcome variable (Y) and the predictor variables ($x_1, x_2, \ldots, x_k$) is linear; (2) The variability of Y, as assessed by the standard deviation σ, corresponding to a particular set of values $x_1, x_2, \ldots, x_k$ is the same, regardless of $x_1, x_2, \ldots, x_k$; (3) The values of the outcome variable Y should have a normal distribution for each set of values of the predictor variables $x_1, x_2, \ldots, x_k$. The assumption of normality of the outcome variable will not be frequently satisfied for HRQoL data. The reason is that many standard HRQoL instruments comprise items on dimensions or domains by addition or weighted average, based on the psychometric properties of the instrument and thus, the measure for each dimension is an ordered categorical (ordinal) scale, as it happens with the SF-36. The distribution of these dimensions could be different, some having relatively few categories and others with a large number of ordered categories, where it is reasonable to assume that there exists an underlying continuous latent variable Z and the actual measured outcome is an ordered scale that reflects contiguous intervals along this continuum. However, these scores tend to be skewed, J-shaped, or even U-shaped. Cox *et al.* discuss the statistical analysis of quality of life assessments in their important contribution to the field [11]. They recommend simplicity of design, analysis and presentation of quality of

DOI: 10.1002/sim

life assessments. Although they concentrate on applications to clinical trials, their methods could be more widely applicable. However, as Olschewski, Schulgen and Schumacher point out in the discussion of the paper by Cox *et al.* [11], the statistical procedures proposed by the authors assume normally distributed HRQoL data. They propose alternatives based on generalized linear models for dealing with the non-normality which often arises in data from HRQoL questionnaires. Lall *et al.* also discuss the use of ordinal regression models for HRQoL scales that are ordinal in nature [15]. Moreover, Walters *et al.* [16] used various ordinal regression models (proportional odds, continuation ratio, polytomous and stereotype) and compared them to conventional statistical methods (t-test and multiple regression) and bootstrap methods to analyse HRQoL data with the SF-36 *role-emotional* domain as the outcome. They concluded that a discrete scale should be treated as continuous if it has seven or more categories and as ordinal otherwise. From a clinical point of view, it is desirable to have standard measurements of the different dimensions of a HRQoL instrument. Therefore, a common technique of statistical analysis is preferable for all the dimensions present in the HRQoL instrument.

We have analysed HRQoL data in patients with eating disorders with the MLR approach. There are two reasons for selecting this specific disease. One is that, even though eating disorders have an important impact on personal and public health [17], there are few studies where the patients' perception of their HRQoL has been published [18]. The other reason is that there are not published studies where predictive factors of HRQoL have been analysed and such information could provide evidence about the effectiveness of different therapeutic options to clinicians [19]. In addition, simulated values of HRQoL by using the normal distribution for the same sample have been analysed with the MLR approach in order to check the robustness of the MLR against normality. The results obtained from real and simulated data have been compared. Differences between both analysis suggest that clinical conclusions obtained from an analysis of HRQoL data with the MLR approach, under non-normality of the response variable, could be incorrect.

Our goal is to select a probability distribution to fit HRQoL data given by the SF-36 questionnaire, so that it could be used as a statistical model for them. The primary purpose of this paper is to use a beta-binomial model to fit HRQoL data given by the SF-36 questionnaire. Thus, beta-binomial regression (BBR) is proposed as a method of analysis of the SF-36 outcome. The performance of this approach is shown in the same sample of patients with eating disorders as before. Simulation values of HRQoL, now by using the beta-binomial distribution, have been created and analysed also with the BBR approach. The results obtained from real and simulated data have been compared in order to show how reasonable the assumption of a beta-binomial distribution for HRQoL data given by the SF-36 questionnaire is. Benefits of the BBR approach with respect to the MLR approach have been discussed based on clinical consequences of conclusions addressed from the results of the data analysis.

The rest of the paper is divided into five sections. Section 2 provides a description of the SF-36 questionnaire for measuring HRQoL. Section 3 focuses on the analysis of a sample of 193 subjects with eating disorders with the MLR approach. It includes a description of the data set, the analysis of the data and the results obtained from the real and the simulated data. Section 4 introduces the beta-binomial model. The application of the proposed methodology to the sample with eating disorders is described in Section 5, with the description of the transformation of the SF-36 scores in order to get a beta-binomial distribution, the assessment of the distribution of HRQoL data to a beta-binomial distribution, the methodology of analysis and the results obtained from the real and the simulated data. Section 6 includes a discussion of the results and some general conclusions.

DOI: 10.1002/sim

2. THE HRQoL QUESTIONNAIRE: SF-36

The SF-36 Health Survey was developed within the Medical Outcomes Study [20]. It measures generic HRQoL concepts relevant across age, disease and treatment groups and it provides a comprehensive, psychometrically sound and efficient way to measure HRQoL from the patients' point of view by scoring standardized responses to standardized question. The SF-36 questionnaire has 36 items, with different answer options. It was constructed to represent eight of the most important health dimensions, which are *physical functioning* (*PF*), *role-physical* (*RP*), *bodily pain* (*BP*), *general health* (*GH*), *vitality* (*VT*), *social functioning* (*SF*), *role-emotional* (*RE*) and *mental health* (*MH*). Each of the eight multi-item scales, which are often called HRQoL domains, contains two to ten items. The first four domains are mainly physical, whereas the last four measure mental aspects of HRQoL. The standardized scoring system is thoroughly described by Ware *et al.* [6]. Scale scores range from 0 to 100, where a higher score indicates a better health status. Thus, the SF-36 generates a profile of HRQoL outcomes on eight dimensions. Two summary measures, one physical and one mental, can be created from the eight main domains. The validity and reliability of this instrument has been broadly tested [21]. In addition, there are also several validation studies of the SF-36 questionnaire that compare it to other health measuring instruments [22, 23] or that were used for different diseases, such as human immunodeficiency virus [24], total hip replacement [25] or chronic obstructive pulmonary disease [26]. Because of this, it is one of the most widely used generic instrument to measure HRQoL.

Multiple linear regression is used in many studies for analysing the effect of explanatory variables on HRQoL measured by the SF-36. The scores are treated as if they were from a continuous distribution and were normally distributed. Actually, it is assumed that there exists an underlying latent variable that measures HRQoL and that the actual measured outcomes are ordered categories that reflect contiguous intervals along this continuum. There are several attempts to try to overcome the lack of normality of the outcome, such as controlling with randomization, using non-parametric univariate methods, analysing changes in HRQoL instead of HRQoL itself, or analysing the two summary measures of the SF-36 questionnaire. Some of the statistical problems of analysing data from the SF-36 questionnaire are discussed by Rose *et al.* [27]. In addition, Walters and Campbell have introduced the use of bootstrap methods for analysing HRQoL outcomes, particularly the SF-36 [28]. The beta distribution has also been used by Cheung *et al.* to model the scores of the eight domains of the SF-36 divided by 100, where covariate effects have been incorporated in the estimation of the parameters in order to show differences in mean and variance of quality of life scores between groups [29]. However, we could not find in the literature any other attempt to use the beta or any other probability distribution to analyse the effect of explanatory variables in the SF-36 as the outcome.

3. DATA ANALYSIS WITH THE MLR APPROACH

3.1. Description of the data set

The sample we have considered is from a longitudinal study of patients diagnosed with an eating disorder who were followed in the Eating Disorders Outpatient Clinic of the Psychiatric Service at Galdakao Hospital, Spain. This work is restricted to the first visit and to a subsample of 193 female patients out of a total of 197. The main goal of the study was to check the influence of

DOI: 10.1002/sim

Table I. Description of the cohort of women with eating disorders at their first visit to the psychiatric service of the hospital.

N	193
Age: $\overline{x}$ (SD)	23.4 (6.7)
Duration of the illness (years): $\overline{x}$ (SD)	6.3 (6.4)
Body mass index: $\overline{x}$ (SD)	19.9 (4.7)
Diagnosis: *n* (%)	
Anorexia Nervosa	54 (28%)
Bulimic Anorexia	60 (31%)
Bulimia Nervosa/Binge Eating	79 (41%)
Severity: *n* (%)	
Mild	56 (29%)
Moderate	54 (28%)
Severe	83 (43%)
Anxiety: *n* (%)	
No	70 (36%)
Yes	123 (64%)
Depression: *n* (%)	
No	136 (70%)
Yes	57 (30%)
SF-36: $\overline{x}$ (SD)	
Physical functioning	87.3 (15.9)
Role-physical	55.3 (42.1)
Bodily pain	66.2 (27.7)
General health	49.6 (21.4)
Vitality	46.7 (22.1)
Social functioning	55.9 (29.5)
Role-emotional	45.7 (43.8)
Mental health	45.4 (22.8)

Abbreviation of standard deviation (SD) is used.

clinical and sociodemographical variables in HRQoL of women with eating disorders. In other words, clinicians wanted to know which of the explanatory variables have a statistically significant influence on the HRQoL reported by the patient and to measure the magnitude of this influence. The variables that we consider are: age; body mass index (BMI); duration of the illness; diagnosis, according to the Diagnostic and Statistical Manual of Mental Disorders, 4th edition [30]; level of anxiety and depression, using the Hospital Anxiety and Depression Scale [31]; severity of the disease, using the Eating Attitudes Test [32] and HRQoL using the SF-36 questionnaire. A sociodemographic, clinical and HRQoL summary of the sample is presented in Table I and a more detailed description of the data set and the way data were collected can be found in Reference [18].

3.2. Method of analysis

Normal fitting of the eight domains of HRQoL was performed using unbiased estimates of the parameters of the normal distribution showed in Table I. Goodness-of-fit to a normal distribution was conducted by the Kolmogorov–Smirnov test.

A clinically important aspect of the study was to analyse the influence of patients' and disease's characteristics on HRQoL. In order to do this, data were analysed using the MLR approach.

DOI: 10.1002/sim

Therefore, the model is given by

$$y_{ij} = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_j + \varepsilon_{ij}, \quad i = 1, \ldots, N, \quad j = 1, \ldots, 8 \tag{1}$$

where y_{ij} represents the response, for subject $i = 1, \ldots, N$ on domain $j = 1, \ldots, 8$; $\mathbf{x}_i$ is a vector of length k (k-vector) of explanatory variables observed on subject i; $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jk})^{\mathrm{T}}$ are k-vectors of unknown regression coefficients on domain $j = 1, \ldots, 8$ and the ε_{ij}'s are uncorrelated, normally distributed random variables with mean zero and constant variance across subjects throughout the range of fitted values for each domain $j = 1, \ldots, 8$, i.e. $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma_j^2)$, representing the deviation of the response from the model prediction $\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_j$, for each $j = 1, \ldots, 8$. Separate analysis are performed for each domain $j = 1, \ldots, 8$, thus variance depends upon the subscript j, representing the domain. Thus, model 1 assumes that the dependent variable, HRQoL score, is continuous and also implies that the dependent variable is normally distributed.

Assumptions for MLR were checked graphically and goodness-of-fit to a normal distribution for the residuals was conducted by the Kolmogorov–Smirnov test.

In order to assess the robustness of MLR to departures from normality, a simulation study was conducted. For the 193 patients with eating disorders in the data set, simulated values of the SF-36 eight domains have been obtained using a normal distribution. The parameters of the normal distribution have been estimated using the unbiased estimates of the mean and the variance showed in Table I. Values of the covariates have been assigned to the simulated values of the eight HRQoL domains. For consistency, proportions of the covariates in the original data set have been maintained also for simulated data. Simulated data were also analysed using the MLR approach.

In both cases, separate analysis were performed for each of the eight HRQoL domains. All of the explanatory variables were tested for significance. Model selection was performed by the F-test for nested models. Significance level was stated at $\alpha = 0.05$.

All the statistical analysis was conducted with SAS System for Windows Version 8.02 [33] and graphical displays were obtained with S-plus 2000 [34].

3.3. Results

Figure 1 shows the histograms for the eight HRQoL domains of the SF-36 questionnaire. The expected curve of the probability density function under the assumption of a normal distribution is overlaid to the histograms in Figure 1. Goodness-of-fit to a normal distribution was rejected for seven out of eight domains of HRQoL ($\alpha = 0.05$). *Vitality* was the only domain where the normal distribution was not rejected ($p = 0.09$).

Normal quantile–quantile plots of the residuals of the finally selected MLR models for the eight domains of the SF-36 questionnaire are shown in Figure 2. They should be close to a straight line if the residuals were normally distributed, which in general are not. Goodness-of-fit to a normal distribution was clearly rejected for the *physical functioning*, *role-physical* and *role-emotional* domains ($p < 0.001$), which are the most skewed areas, and also for *general health* ($p = 0.02$). It was not rejected for the *bodily pain*, *vitality*, *mental health* and *social functioning* domains. Homocedasticity was also checked graphically plotting residuals against predicted values and residuals against covariates (scatter plot for continuous variables and box plots for categorical variables). Scatter plots of residuals against the predicted values do not show violation of the constant variance assumption. However, some decrease in the variance of the residuals was observed, as values of

DOI: 10.1002/sim

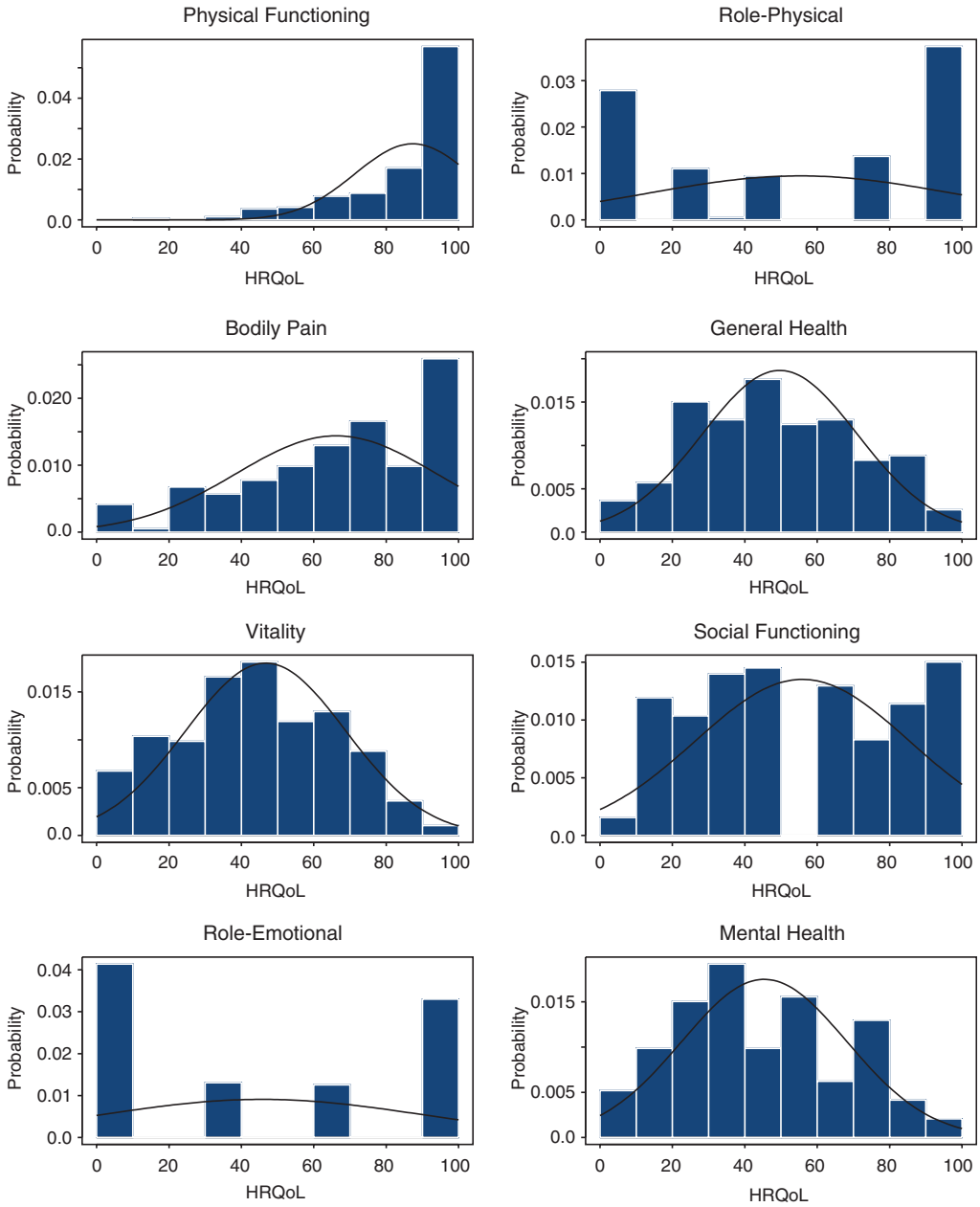

Figure 1. Observed probability (bars) and expected probability under the normal distribution (curve) for the eight domains of the SF-36 questionnaire.

the covariates increase in some of the HRQoL domains. No violation of the linearity assumption was found in the graphical displays of residuals *versus* predictors except for some high values of BMI or duration of the illness.

DOI: 10.1002/sim

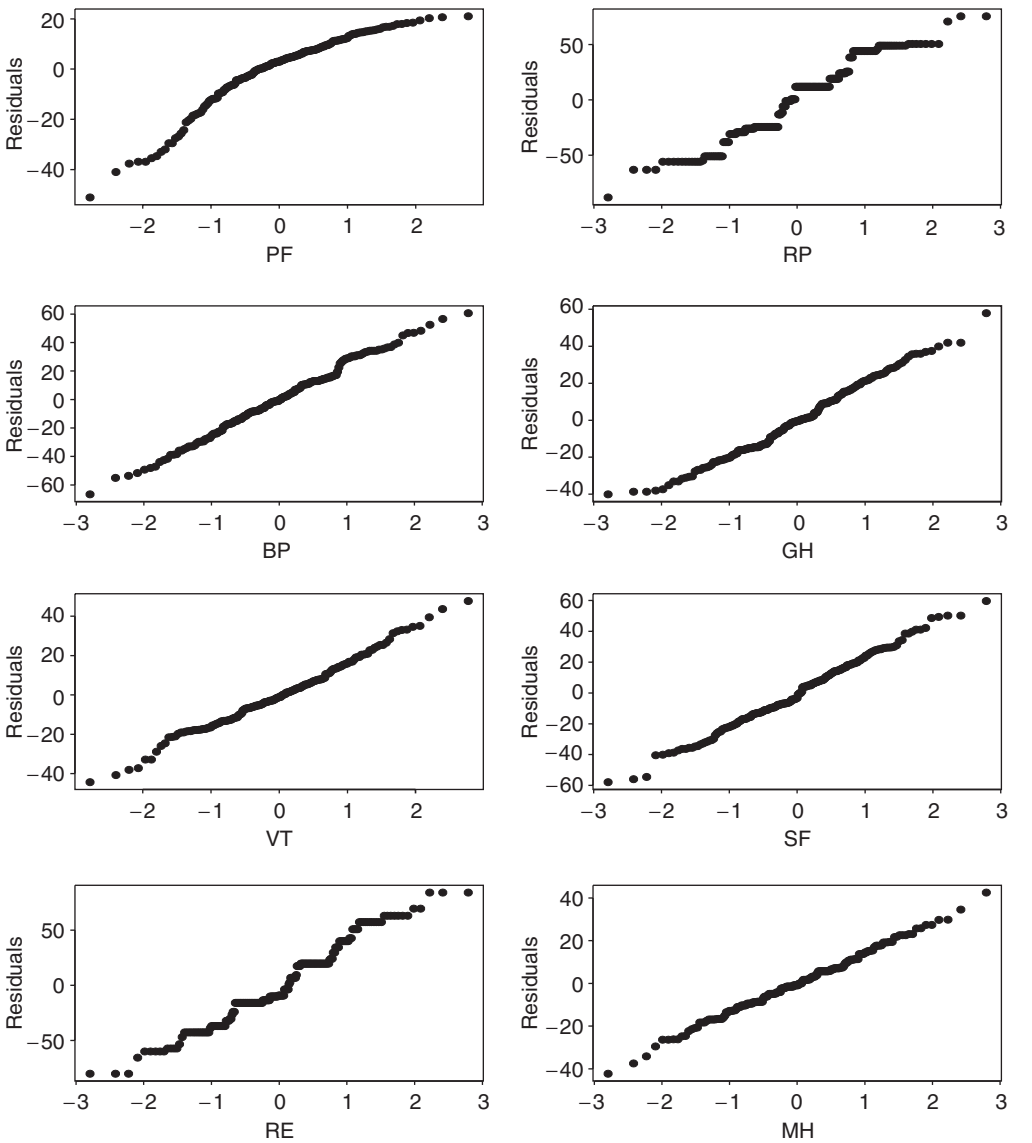

Figure 2. Normal quantile–quantile plot of the residuals of the finally selected multiple linear regression model for the eight domains of the SF-36 questionnaire.

Tables II and III show the estimates of the coefficients, their standard deviations and their statistical significance under the Wald test in the MLR approach with all the covariates in the model for each of the eight real and simulated HRQoL domains divided in physical and mental aspects of HRQoL.

The estimates of the coefficients from both real and simulated data sets are very similar for the *general health* and *vitality* HRQoL domains. The only estimates in these two domains that show differences from real to simulated data are for small values that are never significant in any of

DOI: 10.1002/sim

Table II. Estimates of the coefficients, standard deviations and p-values associated to the multiple linear regression model for each of the four physical HRQoL domains of the SF-36 questionnaire for real and simulated data.

			Real data			Simulated data		
HRQoL	Covariates		$\hat{\beta}$	$\hat{SD}(\hat{\beta})$	p	$\hat{\beta}$	$\hat{SD}(\hat{\beta})$	p
PF	Age		−0.14	0.26	0.5877	−0.09	0.26	0.7466
	Duration†*		−0.35	0.28	0.2046	−0.36	0.28	0.2094
	BMI†*		−1.20	0.31	0.0002	−1.09	0.32	0.0008
	Diagnosis:†*	BA	5.95	2.86	0.0261	7.53	2.93	0.0075
		BN/BE	8.11	3.16		9.32	3.22	
	Severity:†	Moderate	−3.08	2.99	0.477	−5.83	3.06	0.0506
		Severe	−3.55	3.13		−7.63	3.2	
	Anxiety†*		−5.61	2.82	0.0482	−8.88	2.88	0.0024
	Depression†*		−7.37	2.74	0.0078	−8.66	2.8	0.0023
RP	Age		−0.06	0.72	0.9337	−0.02	0.69	0.9779
	Duration†		−0.81	0.75	0.2801	−1.03	0.72	0.1519
	BMI†		−0.90	0.79	0.2571	−1.47	0.76	0.0543
	Diagnosis:†	BA	5.74	7.15	0.6919	12.34	6.86	0.0610
		BN/BE	5.32	7.97		17.25	7.65	
	Severity:†*	Moderate	−26.62	7.57	0.0001	−23.05	7.26	0.0001
		Severe	−33.28	7.93		−32.68	7.61	
	Anxiety†		−4.06	7.06	0.5657	−13.22	6.78	0.0527
	Depression†*		−27.15	6.93	0.0001	−25.31	6.65	0.0002
BP	Age†*		−0.64	0.45	0.1529	−0.34	0.41	0.4174
	Duration		−0.15	0.48	0.7578	−0.33	0.44	0.4558
	BMI		−0.56	0.54	0.2999	−0.30	0.5	0.5509
	Diagnosis:	BA	2.18	4.95	0.8493	4.48	4.55	0.6015
		BN/BE	−0.52	5.45		1.47	5.02	
	Severity:†*	Moderate	−8.13	5.17	0.0162	−7.43	4.76	0.0101
		Severe	−15.72	5.42		−15.25	4.98	
	Anxiety		−4.92	4.88	0.3141	−7.29	4.49	0.1060
	Depression†*		−14.14	4.73	0.0032	−11.24	4.35	0.0106
GH	Age		0.15	0.36	0.6711	0.07	0.27	0.8091
	Duration†*		−0.54	0.38	0.1584	−0.36	0.29	0.2149
	BMI		−0.12	0.43	0.7859	−0.19	0.33	0.5721
	Diagnosis:	BA	4.51	3.93	0.5064	3.26	3.01	0.5571
		BN/BE	1.62	4.33		1.82	3.31	
	Severity:	Moderate	−2.34	4.11	0.2694	−2.04	3.14	0.2080
		Severe	−6.74	4.3		−5.67	3.29	
	Anxiety†*		−11.30	3.87	0.004	−8.49	2.96	0.0047
	Depression		−5.16	3.76	0.1718	−4.32	2.87	0.1340

The following abbreviations are used: physical functioning (PF), role-physical (RP), bodily pain (BP), general health (GH), Bulimic Anorexia (BA), Bulimia Nervosa (BN) and Binge Eating (BE).

Significant covariates in model selection are showed: *real and †simulated.

DOI: 10.1002/sim

Table III. Estimates of the coefficients, standard deviations and p-values associated to the multiple linear regression model for each of the four mental HRQoL domains of the SF-36 questionnaire for real and simulated data.

			Real data			Simulated data		
HRQoL	Covariates		$\hat{\beta}$	$\widehat{\mathrm{SD}}(\hat{\beta})$	p	$\hat{\beta}$	$\widehat{\mathrm{SD}}(\hat{\beta})$	p
VT	Age		−0.06	0.3	0.8355	0.03	0.33	0.9153
	Duration		−0.21	0.33	0.5204	−0.31	0.35	0.3724
	BMI†*		−0.55	0.37	0.1364	−0.59	0.4	0.1404
	Diagnosis:	BA	−1.51	3.37	0.8239	−1.75	3.62	0.8046
		BN/BE	0.56	3.71		0.60	3.99	
	Severity:†*	Moderate	−8.04	3.52	0.0041	−8.81	3.79	0.0033
		Severe	−12.35	3.69		−13.52	3.96	
	Anxiety†*		−11.03	3.32	0.0011	−12.38	3.57	0.0007
	Depression†*		−15.91	3.22	<0.0001	−15.40	3.46	<0.0001
SF	Age†*		1.12	0.42	0.0083	1.49	0.4	0.0003
	Duration†*		−1.40	0.45	0.0022	−1.56	0.43	0.0004
	BMI		−0.45	0.51	0.3808	−0.67	0.49	0.1730
	Diagnosis:	BA	−2.79	4.63	0.7025	−1.11	4.45	0.7676
		BN/BE	1.06	5.11		2.30	4.91	
	Severity:†*	Moderate	−11.75	4.84	0.0159	−11.75	4.66	0.0054
		Severe	−13.84	5.07		−15.47	4.87	
	Anxiety†*		−18.84	4.56	<0.0001	−17.55	4.39	<0.0001
	Depression†*		−16.89	4.48	0.0002	−18.39	4.26	<0.0001
RE	Age†		0.41	0.68	0.5454	1.76	0.61	0.0047
	Duration†		−0.22	0.73	0.7584	−1.61	0.66	0.0153
	BMI		−0.94	0.82	0.2523	−1.05	0.74	0.1566
	Diagnosis:	BA	−6.44	7.5	0.6916	−2.24	6.78	0.9471
		BN/BE	−3.24	8.31		−1.21	7.5	
	Severity:†*	Moderate	−18.28	7.84	0.0685	−21.71	7.08	0.0052
		Severe	−12.08	8.21		−20.66	7.41	
	Anxiety†*		−24.18	7.38	0.0013	−21.57	6.67	0.0015
	Depression†*		−26.00	7.18	0.0004	−27.91	6.48	<0.0001
MH	Age		0.25	0.26	0.3280	0.36	0.26	0.1703
	Duration		−0.21	0.27	0.4393	−0.30	0.28	0.2853
	BMI		−0.59	0.31	0.0595	−0.53	0.32	0.0954
	Diagnosis:	BA	1.29	2.86	0.0630	0.59	2.9	0.1193
		BN/BE	6.99	3.11		6.03	3.19	
	Severity:†*	Moderate	−8.40	2.95	<0.0001	−8.84	3.03	<0.0001
		Severe	−14.69	3.09		−14.66	3.17	
	Anxiety†*		−17.58	2.78	<0.0001	−17.35	2.85	
	Depression†*		−17.13	2.7	<0.0001	−17.48	2.77	<0.0001

The following abbreviations are used: vitality (VT), social functioning (SF), role-emotional (RE), mental health (MH), Bulimic Anorexia (BA), Bulimia Nervosa (BN) and Binge Eating (BE).

Significant covariates in model selection are showed: *real and †simulated.

DOI: 10.1002/sim

these models. There are some differences between estimates of the coefficients of the simulated and the real data for the *bodily pain*, *social functioning* and *mental health* domains. However, these differences do not affect significant variables in the model except for the *mental health* domain, where the coefficient for diagnosis (Bulimic Anorexia) for real data is more than twice than for simulated data. Finally, differences between estimates of the coefficients of the simulated and real data are considerably large for the *physical functioning*, *role-physical* and *role-emotional* domains. For *physical functioning*, estimates came out to be quite different for severity (-3.08 vs -5.83 and -3.55 vs -7.63), they turned out to be larger in magnitude for simulated than for real data. For *role-physical* estimates came out to be extremely different for diagnosis (5.74 vs 12.34 and 5.32 vs 17.25) and for anxiety (-4.06 vs -13.22), they were both larger in magnitude for simulated than for real data. For *role-emotional* some of the estimates came out very different for duration of the illness (-0.22 vs -1.61) and for severity (-18.28 vs -21.71 and -12.08 vs -20.66), they were both larger in magnitude for simulated than for real data, and for diagnosis (-6.44 vs -2.24 and -3.24 vs -1.21), they turned out to be larger in magnitude for real than for simulated data.

The estimated standard deviations of the estimates of the coefficients are very similar for both real and simulated data sets. The ratio of the standard deviations of the coefficients obtained from the real data and the simulated data (largest/smallest) is between 1.02 (*physical functioning* and *mental health*) and 1.31 (*general health*).

The significance of the covariates in the model that includes all of them was very similar for real and simulated data in the following HRQoL domains: *general health* and *vitality*. There were minor differences for *bodily pain*, *social functioning* and *mental health*. However, there were important differences in some p-values between real and simulated data for the remaining three HRQoL domains. For *physical functioning* the p-value for severity was 0.48 for real data vs 0.05 for simulated data. For *role-physical* there were differences between p-values of BMI (0.26 vs 0.05), diagnosis (0.69 vs 0.06) and anxiety (0.57 vs 0.05), they turned out to be always larger for real than for simulated data. Also for *role-emotional* there were differences between p-values of age (0.54 vs 0.005) and duration of the illness (0.76 vs 0.02), they were again larger for real than for simulated data.

Model selection has provided the same covariates as significant for real and simulated data in five of the eight HRQoL domains, namely *bodily pain* (age, depression and severity), *general health* (duration of the illness and anxiety), *vitality* (BMI, severity, anxiety and depression), *social functioning* (age, duration of the illness, severity, anxiety and depression) and *mental health* (severity, anxiety and depression). Whereas for *physical functioning*, *role-physical*, and *role-emotional* results from the original data are different from those from the simulated data in the model selection process. Significant covariates in both data sets, real and simulated, for *physical functioning* are BMI, duration of the illness, diagnosis, anxiety and depression; severity is also significant in simulated data set but is not significant in the real data set. For *role-physical*, significant covariates in both data sets are just severity and depression whereas BMI, duration of the illness, diagnosis and anxiety are significant only in the simulated data set. For *role-emotional* severity, anxiety and depression are significant in both cases, but age and duration of the illness are also significant in the simulated data set. As we have mentioned above, these three are also the HRQoL domains where the normality of the residuals in the final models is clearly rejected.

Comparison between results obtained from real and simulated data by using the MLR approach showed that differences are very sensible to the shape of the outcome. There are minor differences when the distribution of the HRQoL domain is not very skewed, like *general health* and *vitality*, moderate when it is J-shaped, like *physical functioning*, and large when it is U-shaped like *role-*

DOI: 10.1002/sim

physical and *role-emotional*. Differences, when they exist, are in the significance of a covariate in the model and in the magnitude of the effect of such covariate in HRQoL. First, selected models with simulated data took more covariates as significant than the ones with real data, which means that, if the MLR approach is used in order to find the effect of some clinical and sociodemographic variables on HRQoL, some important variables could not be included in the model. Second, the magnitude of the effect of the covariates in HRQoL changes in any direction from real to simulated data. For instance, estimates of the coefficients of severity in *physical functioning* and estimates of the coefficients of diagnosis and anxiety in *role-physical* are larger in simulated than in real data. This means that the effect of severity in *physical functioning* and the effect of diagnosis and anxiety in *role-physical* are underestimated when these HRQoL are supposed to be normally distributed and they are not. However, estimates of the coefficients of diagnosis in *role-emotional* are smaller in simulated than in real data. This means that the effect of diagnosis in *role-emotional* is overestimated when the MLR is used. Therefore, we conclude that data analysis of HRQoL with the MLR approach, under non normality of the response variable, does not support clinical conclusions obtained from it. We propose the beta-binomial distribution to fit HRQoL outcome given by the SF-36 questionnaire. In the remaining sections we describe the beta-binomial distribution approach and we perform the data analysis of HRQoL data under this assumption.

4. THE BETA-BINOMIAL MODEL

Consider N independent random variables $Y_1, \ldots, Y_N$ representing the number of successes in n_i binomial trials, $i = 1, \ldots, N$. If θ_i is the true probability of success in each set of trials, the conditional distribution of Y_i given θ_i is binomial with parameters n_i and θ_i. In addition, it is assumed that the θ_i's are a random sample from a beta probability distribution with parameters p_i/ϕ_i and $(1 - p_i)/\phi_i$, $0<p_i<1$, $\phi_i>0$, $i = 1, \ldots, N$. In this setting, it is well known that its mean is p_i and its variance is $p_i(1 - p_i)\phi_i/(1 + \phi_i)$. Depending on the values of its parameters, this beta distribution can be bell-shaped, U-shaped, J-shaped or reverse J-shaped.

Under these conditions, we can derive the marginal distribution of the observed number of successes Y_i, which is called the beta-binomial distribution and has probability density function

$$
\begin{aligned}
f(y_i) &= \int_0^1 P(Y_i = y_i|\theta_i)\, f(\theta_i)\, \mathrm{d}\theta_i \\
&= \int_0^1 \left[\binom{n_i}{y_i} \theta_i^{y_i} (1-\theta_i)^{n_i - y_i} \right] \left[B\left(\frac{p_i}{\phi_i}, \frac{1-p_i}{\phi_i}\right) \theta_i^{(p_i/\phi_i)-1} (1-\theta_i)^{\frac{1-p_i}{\phi_i}-1} \right] \mathrm{d}\theta_i \\
&= \binom{n_i}{y_i} \frac{\Gamma\left(\frac{1}{\phi_i}\right)}{\Gamma\left(\frac{1}{\phi_i} + n_i\right)} \frac{\Gamma\left(\frac{p_i}{\phi_i} + y_i\right)}{\Gamma\left(\frac{p_i}{\phi_i}\right)} \frac{\Gamma\left(\frac{1-p_i}{\phi_i} + n_i - y_i\right)}{\Gamma\left(\frac{1-p_i}{\phi_i}\right)} \\
&= \binom{n_i}{y_i} \frac{\prod_{k=0}^{y_i} (p_i + k\phi_i) \prod_{k=0}^{n_i - y_i} (1 - p_i + k\phi_i)}{\prod_{k=0}^{n_i} (1 + k\phi_i)} \qquad (2)
\end{aligned}
$$

DOI: 10.1002/sim

The use of the beta-binomial distribution to describe the variation in the probability parameter of a binomial distribution was originally motivated and proposed by Skellam and has been widely used since then [35]. The model has been employed, for example, in modelling consumer purchasing behaviour [36], in dental studies of caries in children [37] or in describing disease incidence in households [38]. Its use was also suggested for toxicological data by Williams [39].

Under this model, the mean and the variance of Y_i are given by

$$\begin{aligned} E(Y_i) &= n_i\, p_i \\ \mathrm{Var}(Y_i) &= n_i\, p_i(1-p_i)\left[1+(n_i-1)\frac{\phi_i}{1+\phi_i}\right] \end{aligned} \tag{3}$$

The parameter $\phi_i > 0$ is called the overdispersion parameter. The term $\phi_i/(1+\phi_i) \in [0,1]$ is a correlation-like measure of propagation. When $\phi_i = 0$, it corresponds to the binomial case, meaning that there is no propagation of the variance. Therefore, the factor $[1+(n_i-1)\phi_i/(1+\phi_i)]$ is a multiplier of the binomial variance.

This model can be fitted using the method of maximum likelihood, as described by Williams [39] and Crowder [40]. More specifically, it can be seen from (2) that the log-likelihood is given by

$$\begin{aligned} L(n_i, p_i, \phi_i | y_i) = \sum_{i=1}^{N}\left[\ln\binom{n_i}{y_i} + \sum_{k=0}^{y_i}\ln(p_i+k\phi_i) + \sum_{k=0}^{n_i-y_i}\ln(1-p_i+k\phi_i)\right] \\ -\sum_{i=1}^{N}\sum_{k=0}^{n_i}\ln(1+k\phi_i) \end{aligned} \tag{4}$$

Maximum-likelihood estimates of p_i and ϕ_i require numerical iteration and, thus, the Newton–Raphson method can be used.

5. DATA ANALYSIS UNDER THE BETA-BINOMIAL DISTRIBUTION

5.1. Assessment of the distribution of the HRQoL data

The original HRQoL domains have been recoded as discrete variables using the empirically suggested values. Let Y_{ij} denote the score of the ith individual on the jth domain of HRQoL, $i = 1, \ldots, N$, $j = 1, \ldots, 8$. The recoding process has been done following two steps. First, take the values of the N individuals for the jth domain and order them from smallest to largest. Second, if the number of different values in the ordered list is small, recode them from 0 to n_j. On the other hand, if the number of values in the ordered list is large, categorize the original variable and recode the categories from 0 to n_j. Let Y_{ij}^* denote the recoded variable, it represents the points of the ith individual in the jth domain of HRQoL, $i = 1, \ldots, N$, $j = 1, \ldots, 8$ and it is a discrete variable that takes integer values from 0 to n_j. Summarizing, the eight continuous scales of the SF-36 have been transformed into eight ordinal scales. Table IV shows the relation between the original and the recoded HRQoL variables.

Figure 3 shows the histograms for the eight recoded HRQoL domains of the SF-36 questionnaire. Beta-binomial fitting of the eight domains of HRQoL was performed estimating the parameters

DOI: 10.1002/sim

Table IV. Original (Y_j) and recoded (Y_j^*) HRQoL data for the eight domains of the SF-36 questionnaire.

HRQoL	Y_3	Y_3^*	HRQoL	Y_6	Y_6^*
Bodily pain	[0, 5]	0	Social	[0, 6.25]	0
	(5, 15]	1	functioning	(6.25, 18.75]	1
	(15, 25]	2		(18.75, 31.25]	2
	(25, 35]	3		(31.25, 43.75]	3
	(35, 45]	4		(43.75, 56.25]	4
	(45, 55]	5		(56.25, 68.75]	5
	(55, 65]	6		(68.75, 81.25]	6
	(65, 75]	7		(81.25, 93.75]	7
	(75, 85]	8		(93.75, 100]	8
	(85, 95]	9			
	(95, 100]	10			
HRQoL	Y_2	Y_2^*	HRQoL	Y_7	Y_7^*
Role-physical	[0, 12.5]	0	Role-	[0, 16.67]	0
	(12.5, 37.5]	1	emotional	(16.67, 50]	1
	(37.5, 62.5]	2		(50, 83.33]	2
	(62.5, 87.5]	3		(83.33, 100]	3
	(87.5, 100]	4			
HRQoL	Y_j	Y_j^*		Y_j	Y_j^*
Physical	[0, 2.5]	0		(57.5, 62.5]	12
functioning	(2.5, 7.5]	1		(62.5, 67.5]	13
($j=1$)	(7.5, 12.5]	2		(67.5, 72.5]	14
	(12.5, 17.5]	3		(72.5, 77.5]	15
General health	(17.5, 22.5]	4		(77.5, 82.5]	16
($j=4$)	(22.5, 27.5]	5		(82.5, 87.5]	17
	(27.5, 32.5]	6		(87.5, 92.5]	18
Vitality	(32.5, 37.5]	7		(92.5, 97.5]	19
($j=5$)	(37.5, 42.5]	8		(97.5, 100]	20
	(42.5, 47.5]	9			
Mental health	(47.5, 52.5]	10			
($j=8$)	(52.5, 57.5]	11			

of the distribution by maximum likelihood. Table V shows the estimation of the parameters for the beta-binomial distribution for each of the eight HRQoL domains. The expected curve of the probability density function under the assumption of a beta-binomial distribution is overlaid to the histograms in Figure 3. Goodness-of-fit to a beta-binomial distribution, conducted by χ^2 tests was rejected only for three of the eight domains of HRQoL, namely *bodily pain*, *social functioning* and *mental health* ($\alpha = 0.05$).

5.2. *Method of analysis*

Beta-binomial regression, as a particular case of logistic regression with random effects considering a beta-binomial marginal distribution, was used in order to analyse the influence of patients' and

DOI: 10.1002/sim

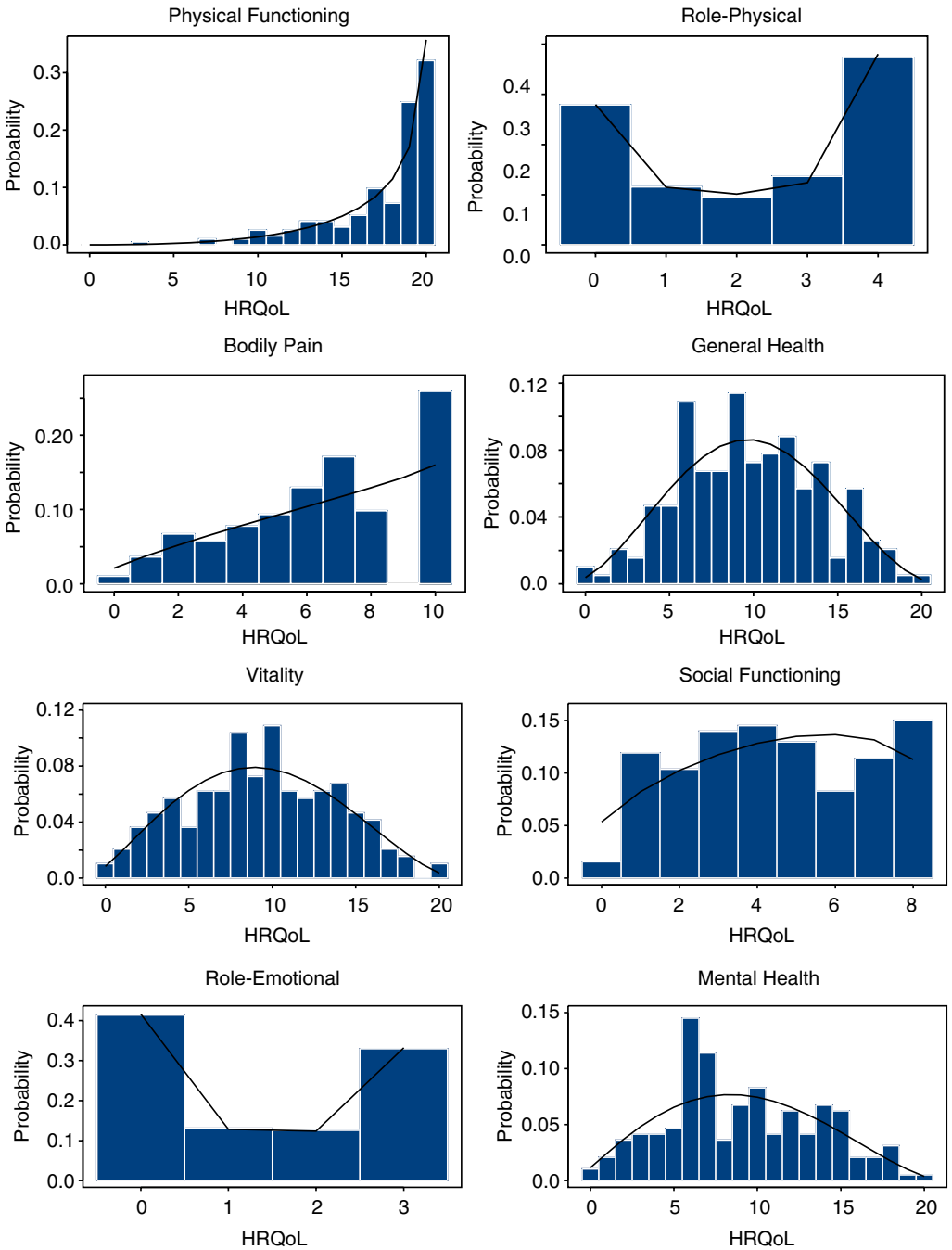

Figure 3. Observed probability (bars) and expected probability under the beta-binomial distribution (curve) for the eight domains of the SF-36 questionnaire.

DOI: 10.1002/sim

Table V. Estimation of the parameters of the BB(n, p, ϕ) distribution.

HRQoL	n	$\hat{p}$	$\hat{\phi}$
Physical functioning	20	0.8726	0.2174
Role-physical	4	0.5512	1.5963
Bodily pain	10	0.5865	0.2243
General health	20	0.4883	0.1487
Vitality	20	0.4663	0.1849
Social functioning	8	0.5671	0.3549
Role-emotional	3	0.4573	1.9021
Mental health	20	0.4551	0.1944

disease's characteristics on HRQoL [41]. Therefore, the model is given by

$$\text{logit}(\theta_{ij}) = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j + \sigma_j u_i, \quad i = 1, \ldots, N, \quad j = 1, \ldots, 8 \tag{5}$$

where as described in Section 4, θ_{ij} represents the true probability of success in each set of n_j binomial trials for the subject $i = 1, \ldots, N$. In this particular case, θ_{ij} represents the probability of obtaining one point on the jth HRQoL domain for subject $i = 1, \ldots, N$. As in model (1), $\mathbf{x}_i$ is a k-vector of explanatory variables observed on subject i and $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jk})^{\mathrm{T}}$ are k-vectors of unknown regression coefficients on the HRQoL domain $j = 1, \ldots, 8$. The constant σ_j is the unknown positive parameter for the random effect on the jth HRQoL domain and the u_i's are independent and identically distributed random variables with mean zero and variance one.

Simulation of the SF-36 eight domains has been repeated using the beta-binomial distribution. Estimates of p and ϕ have been obtained by the maximum-likelihood method and are shown in Table V. Values of the covariates have been assigned to the simulated values as mentioned before. Simulated data were also analysed using the BBR approach.

In both cases, separate analysis were performed for each of the eight HRQoL domains. All of the explanatory variables were tested for significance. Model selection was performed using the χ^2 test for difference in deviances of two nested models [42]. Significance level was stated at $\alpha = 0.05$.

All the statistical analysis was conducted with Egret for Windows Version 2.0.1 [43] and graphical displays were obtained with S-plus 2000 [34].

5.3. Results

Tables VI and VII show the estimates of the coefficients, their standard deviations and their statistical significance in the BBR model with all the covariates in the model for each of the eight real and simulated HRQoL domains divided in physical and mental aspects of HRQoL.

The estimates of the coefficients from both real and simulated data sets are very similar for all the eight HRQoL domains, except for *bodily pain*. In most cases, the estimates that show differences from real to simulated data are very small quantities that are never significant in any of these models. For *bodily pain*, the estimate of the coefficient of age is larger in magnitude for real than for simulated data (-0.03 vs -0.01), whereas the opposite happens for the estimate of the coefficient of duration of the illness (-0.005 vs -0.02). The estimated standard deviations of the estimates of the coefficients from real and simulated data are also very similar. The ratio of the standard deviations of the coefficients obtained from the real data and the simulated data

DOI: 10.1002/sim

Table VI. Estimates of the coefficients, standard deviations and p-values associated to the beta-binomial regression model for each of the four physical HRQoL domains of the SF-36 questionnaire for real and simulated data.

			Real data			Simulated data		
HRQoL	Covariates		$\hat{\beta}$	$\widehat{\mathrm{SD}}(\hat{\beta})$	p	$\hat{\beta}$	$\widehat{\mathrm{SD}}(\hat{\beta})$	p
PF	Age		−0.01	0.02	0.5972	−0.004	0.02	0.8638
	Duration†*		−0.02	0.02	0.2723	−0.03	0.02	0.1696
	BMI†*		−0.09	0.02	<0.001	−0.08	0.02	<0.001
	Diagnosis:†*	BA	0.65	0.24	0.0115	0.70	0.25	0.0084
		BN/BE	0.65	0.26		0.66	0.27	
	Severity:	Moderate	−0.43	0.26	0.1796	−0.47	0.27	0.1258
		Severe	−0.46	0.26		−0.53	0.27	
	Anxiety†*		−0.45	0.25	0.0666	−0.55	0.26	0.0311
	Depression†*		−0.60	0.21	0.0043	−0.62	0.21	0.0035
	Scale parameter ϕ		0.1417	0.0254		0.155	0.0276	
RP	Age		−0.01	0.03	0.7483	−0.01	0.03	0.8495
	Duration†*		−0.04	0.04	0.2909	−0.04	0.03	0.3015
	BMI		−0.03	0.04	0.4406	−0.06	0.04	0.1545
	Diagnosis:	BA	0.20	0.36	0.8387	0.29	0.36	0.6620
		BN/BE	0.19	0.4		0.30	0.40	
	Severity:†*	Moderate	−1.57	0.4	<0.001	−1.65	0.39	<0.001
		Severe	−1.73	0.41		−1.78	0.40	
	Anxiety		−0.25	0.35	0.471	−0.34	0.34	0.3228
	Depression†*		−1.32	0.34	<0.001	−1.31	0.33	<0.001
	Scale parameter ϕ		1.0109	0.1943		0.8857	0.1726	
BP	Age*		−0.03	0.02	0.1303	−0.01	0.02	0.4182
	Duration†		−0.005	0.02	0.8215	−0.02	0.02	0.3676
	BMI		−0.02	0.02	0.3384	−0.02	0.02	0.3919
	Diagnosis:	BA	0.18	0.24	0.7314	0.20	0.19	0.5320
		BN/BE	0.03	0.27		0.02	0.21	
	Severity:†*	Moderate	−0.45	0.26	0.0160	−0.29	0.20	0.0047
		Severe	−0.76	0.26		−0.67	0.21	
	Anxiety		−0.27	0.24	0.2628	−0.28	0.19	0.1391
	Depression†*		−0.58	0.22	0.0088	−0.50	0.18	0.0055
	Scale parameter ϕ		0.2918	0.0477		0.1398	0.0271	
GH	Age		0.01	0.01	0.6075	0.01	0.01	0.6957
	Duration†*		−0.02	0.01	0.1205	−0.02	0.01	0.1442
	BMI		−0.005	0.02	0.7645	−0.006	0.02	0.7259
	Diagnosis:	BA	0.18	0.15	0.4773	0.18	0.14	0.4390
		BN/BE	0.08	0.17		0.07	0.16	
	Severity:	Moderate	−0.10	0.16	0.2003	−0.14	0.15	0.1279
		Severe	−0.29	0.17		−0.31	0.16	
	Anxiety†*		−0.43	0.15	0.0042	−0.42	0.14	0.0028
	Depression		−0.20	0.15	0.1729	−0.20	0.14	0.1545
	Scale parameter ϕ		0.1057	0.0166		0.0852	0.0144	

The following abbreviations are used: physical functioning (PF), role-physical (RP), bodily pain (BP), general health (GH), Bulimic Anorexia (BA), Bulimia Nervosa (BN) and Binge Eating (BE).
Significant covariates in model selection are showed: *real and †simulated.

DOI: 10.1002/sim

Table VII. Estimates of the coefficients, standard deviations and p-values associated to the beta-binomial regression model for each of the four mental HRQoL domains of the SF-36 questionnaire for real and simulated data.

		Real data			Simulated data		
HRQoL	Covariates	$\hat{\beta}$	$\widehat{\mathrm{SD}}(\hat{\beta})$	p	$\hat{\beta}$	$\widehat{\mathrm{SD}}(\hat{\beta})$	p
VT	Age	−0.001	0.01	0.9716	−0.002	0.01	0.8939
	Duration	−0.01	0.01	0.4555	−0.01	0.01	0.4449
	BMI†*	−0.02	0.02	0.1479	−0.02	0.01	0.2080
	Diagnosis: BA	−0.08	0.14	0.7401	−0.05	0.14	0.8923
	BN/BE	0.03	0.16		0.01	0.15	
	Severity:†* Moderate	−0.33	0.15	0.0032	−0.32	0.14	0.0038
	Severe	−0.52	0.15		−0.48	0.15	
	Anxiety†*	−0.46	0.14	<0.001	−0.47	0.13	<0.001
	Depression†*	−0.67	0.14	<0.001	−0.65	0.13	<0.001
	Scale parameter ϕ	0.0763	0.0138		0.0618	0.0122	
SF	Age†*	0.06	0.02	0.0041	0.050	0.02	0.0078
	Duration†*	−0.07	0.02	<0.001	−0.07	0.02	<0.001
	BMI	−0.02	0.02	0.2955	−0.02	0.02	0.3543
	Diagnosis: BA	−0.10	0.21	0.7723	−0.06	0.22	0.8556
	BN/BE	0.05	0.24		0.07	0.24	
	Severity:†* Moderate	−0.57	0.23	0.0108	−0.70	0.23	0.0014
	Severe	−0.67	0.23		−0.81	0.24	
	Anxiety†*	−0.88	0.21	<0.001	−0.90	0.21	<0.001
	Depression†*	−0.72	0.19	<0.001	−0.81	0.20	<0.001
	Scale parameter ϕ	0.1512	0.0322		0.1548	0.0329	
RE	Age	0.03	0.03	0.3538	0.04	0.04	0.2114
	Duration	−0.02	0.04	0.6743	−0.02	0.04	0.5079
	BMI	−0.05	0.04	0.1985	−0.05	0.04	0.2385
	Diagnosis: BA	−0.22	0.37	0.8367	−0.24	0.38	0.7480
	BN/BE	−0.14	0.41		0.03	0.42	
	Severity:†* Moderate	−0.91	0.38	0.0552	−0.99	0.38	0.0316
	Severe	−0.59	0.40		−0.69	0.4	
	Anxiety†*	−0.98	0.35	0.0046	−0.98	0.35	0.0049
	Depression†*	−1.34	0.37	<0.001	−1.59	0.40	<0.001
	Scale parameter ϕ	1.1085	0.2466		1.1266	0.2578	
MH	Age	0.02	0.01	0.1688	0.01	0.01	0.1758
	Duration	−0.01	0.01	0.2832	−0.01	0.01	0.4471
	BMI	−0.03	0.01	0.0467	−0.02	0.01	0.1095
	Diagnosis: BA	0.05	0.13	0.0411	0.05	0.12	0.1296
	BN/BE	0.33	0.14		0.25	0.13	
	Severity:†* Moderate	−0.36	0.13	<0.001	−0.35	0.12	<0.001
	Severe	−0.64	0.13		−0.57	0.13	
	Anxiety†*	−0.74	0.12	<0.001	−0.63	0.11	<0.001
	Depression†*	−0.78	0.12	<0.001	−0.86	0.12	<0.001
	Scale parameter ϕ	0.0373	0.0095		0.0292	0.0086	

The following abbreviations are used: vitality (VT), social functioning (SF), role-emotional (RE), mental health (MH), Bulimic Anorexia (BA), Bulimia Nervosa (BN) and Binge Eating (BE).
Significant covariates in model selection are showed: *real and †simulated.

DOI: 10.1002/sim

(largest/smallest) is between 1.002 (*role-emotional* and *mental health*) and 1.27 (*bodily pain*). The significance of the covariates in the model that includes all of them is quite similar for real and simulated data for most of the HRQoL domains. The only domains where there are some differences in the p-values are *physical functioning*, where anxiety is significant for simulated data and it is not for real data (0.067 vs 0.031), and *mental health* where diagnosis and BMI are significant for real data whereas they are not for simulated data (0.0411 vs 0.1296 and 0.0467 vs 0.1095). Model selection has given the same covariates as significant for real and simulated data in all of the HRQoL domains except for *bodily pain*. For *bodily pain*, severity and depression were significant for real and simulated data; but age was significant only for real data and duration of the illness was significant only for simulated data. This difference in model selection could be due to the existing correlation between age and duration of the illness.

Estimates of the scale parameter from real and simulated data are quite similar for all the HRQoL domains, except for *bodily pain* (0.29 vs 0.14).

There are very few differences between the results obtained from real and simulated data when the BBR model is used. Besides, these differences are basically in *bodily pain* and *mental health*, which are two of the three HRQoL domains that do not fit well to a beta-binomial distribution. *Bodily pain* is also the domain with the largest differences between the estimates of the standard deviations of the coefficients, with the ones in the real data set larger than the ones in the simulated data set.

If the aim of the study is to detect statistically significant relationships between covariates and HRQoL, we see how MLR and BBR lead to the same variables as significant predictors of HRQoL for seven out of eight domains. The only exception is *role-physical*, where duration of the illness is significant for BBR ($p = 0.005$) whereas it is not for MLR. Researchers in the area agree that duration of the illness is supposed to be an important predictor of HRQoL, especially in physical and functional aspects, as it is for *physical functioning*, *general health* and *social functioning*.

However, there are more differences between the two approaches when the application is focused on the magnitude of the relationship between the covariates and HRQoL and the way that this relationship is interpreted from a clinical point of view. Final results for MLR and BBR, after model selection process, are shown in Tables VIII and IX, respectively. Let us focus on the interpretation of a particular coefficient in both models, MLR and BBR, as an illustrative example. For *role-physical* and the covariate depression, MLR shows a coefficient of -26.64, whose interpretation is that a patient with depression is expected to have 27 points less (out of 100) in his/her HRQoL than a patient without depression, adjusted by severity. The BBR approach for the same outcome and covariate shows a coefficient of -1.3954, which exponentiated is 0.2478 and its inverse is 4.04. This means that a patient without depression is four times more likely to have a one point (out of four) better *role-physical* than a patient with depression, adjusted by severity and duration of the illness. As results for the outcome are more easily interpreted in terms of the entire scale, from 0 to 100, which is the same for the eight domains, we reverse the recoding process in order to have a relative interpretation of the expression 'one point' of HRQoL. As a consequence of the recodification process detailed in Section 5.1 and results in Table IV, we can consider that one point in the recoded *role-physical* scale is equivalent to 25 points in the original scale, which is a fourth of the entire scale. Therefore, the result from the BBR model can also be interpreted as follows: the presence of depression decreases to 25 per cent the likelihood of having 25 points better *role-physical*, adjusted by severity and duration of the illness. This is a statement that is much more meaningful from a clinical point of view (yet still scientifically responsible) than the expected 27 points change in the HRQoL domain or any p-value. As an additional example we

DOI: 10.1002/sim

Table VIII. Estimates of the coefficients of the final models that include only the significant covariates with multiple linear regression for each of the HRQoL domains of the SF-36 questionnaire.

HRQoL	Covariates	$\hat{\beta}$	HRQoL	Covariates	$\hat{\beta}$
PF	Intercept	115.94	VT	Intercept	79.12
	Duration	−0.4882		BMI	−0.6486
	BMI	−1.1873		Severity: Moderate	−8.43
	Diagnosis: BA	5.40		Severe	−13.09
	BN/BE	7.78		Anxiety	−10.67
	Anxiety	−7.16		Depression	−16.12
	Depression	−7.69			
RP	Intercept	88.27	SF	Intercept	68.02
	Severity: Moderate	−32.33		Age	0.9957
	Severe	−37.16		Duration	−1.4105
	Depression	−26.64		Severity: Moderate	−11.24
				Severe	−15.61
				Anxiety	−19.22
				Depression	−15.66
BP	Intercept	100.99	RE	Intercept	80.27
	Age	−0.8223		Severity: Moderate	−20.33
	Severity: Moderate	−12.33		Severe	−14.69
	Severe	−17.86		Anxiety	−22.92
	Depression	−14.96		Depression	−26.79
GH	Intercept	63.37	MH	Intercept	70.31
	Duration	−0.4978		Severity: Moderate	−8.14
	Anxiety	−16.77		Severe	−15.29
				Anxiety	−17.55
				Depression	−16.58

The following abbreviations are used: physical functioning (PF), role-physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role-emotional (RE), mental health (MH), Bulimic Anorexia (BA), Bulimia Nervosa (BN) and Binge Eating (BE).

can also indicate that a moderate or severe patient is 5.3 or 6.3 times more likely to have a 25 points worse *role-physical* than a mild patient, respectively, adjusted by depression and duration of the illness. Results are also clinically very meaningful in the *role-emotional* domain. For instance, a patient without depression is four times more likely to have 33 points (a third of the entire scale) better *role-emotional* than a patient with depression, adjusted by severity and anxiety, whereas a patient without anxiety is 2.5 times more likely to have 33 points better *role-emotional* than a patient with anxiety, adjusted by severity and depression.

6. DISCUSSION

Our work has concentrated primarily on observational studies where HRQoL has been measured as the main outcome. Therefore, our conclusions focus on the two main purposes of this kind of applications. The first one is to detect statistically significant predictors of HRQoL and the second is to measure the clinical significance of these relationships.

DOI: 10.1002/sim

Table IX. Estimates of the coefficients of the final models that include only the significant covariates with beta-binomial regression for each of the HRQoL domains of the SF-36 questionnaire.

HRQoL	Covariates	$\hat{\beta}$	HRQoL	Covariates	$\hat{\beta}$
PF	Intercept	4.0868	VT	Intercept	1.1798
	Duration	−0.0329		BMI	−0.0259
	BMI	−0.0822		Severity: Moderate	−0.3402
	Diagnosis: BA	0.5986		Severe	−0.5540
	BN/BE	0.5819		Anxiety	−0.4439
	Anxiety	−0.6467		Depression	−0.6750
	Depression	−0.6152			
RP	Intercept	2.3137	SF	Intercept	0.8263
	Duration	−0.0543		Age	0.0496
	Severity: Moderate	−1.6647		Duration	−0.0691
	Severe	−1.8418		Severity: Moderate	−0.5554
	Depression	−1.3954		Severe	−0.7359
				Anxiety	−0.9003
				Depression	−0.6673
BP	Intercept	2.3066	RE	Intercept	1.3044
	Age	−0.0380		Severity: Moderate	−0.9612
	Severity: Moderate	−0.6451		Severe	−0.6733
	Severe	−0.8393		Anxiety	−0.9002
	Depression	−0.6262		Depression	−1.3833
GH	Intercept	0.4970	MH	Intercept	0.8621
	Duration	−0.0210		Severity: Moderate	−0.3478
	Anxiety	−0.6567		Severe	−0.6618
				Anxiety	−0.7305
				Depression	−0.7525

The following abbreviations are used: physical functioning (PF), role-physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role-emotional (RE), mental health (MH), Bulimic Anorexia (BA), Bulimia Nervosa (BN) and Binge Eating (BE).

The decision of comparing the proposed BBR methodology only to the MLR approach and not to other approaches, as ordinal models or bootstrap methods, is because MLR is the most commonly used technique to analyse HRQoL outcomes, particularly the SF-36, in the literature. Criticism of benefits or limitations of techniques other than MLR and BBR is not the purpose of this paper, although they will be mentioned when it applies.

First of all, in order to detect statistically significant predictors of HRQoL, we conclude that the selection of the method of analysis might be done based on the discreteness of the probability distribution of the outcome. When the distribution of the HRQoL domain has a limited number of discrete values, like *role-physical* and *role-emotional*, MLR and BBR provide quite different results. Results from the BBR method were similar to the results obtained by simulation, whereas the MLR results differ also from real to simulated data. Besides, the BBR approach is more powerful for detecting significant relationships, as we have seen for *role-physical*. In addition, the categorization process is quite natural for these domains, because they are discrete variables in nature. Therefore, the BBR approach is preferred to MLR to analyse this kind of HRQoL domains. Ordinal regression methods (proportional odds or partial proportional odds), as suggested

DOI: 10.1002/sim

by other authors [15, 16], could also be selected as methods of analysis. In our experience, results from proportional odds model are very similar to BBR results when the proportional odds assumption is acceptable. However, when partial proportional odds need to be incorporated for some of the covariates, interpretation of the parameter estimates is cumbersome from a clinical point of view.

When the HRQoL domain has a large number of ordered categories, most of which are occupied by the data, the underlying scale can clearly be considered continuous. In addition, if the shape is close to a normal distribution, or at least it is a symmetrical distribution with not very heavy tails, like *general health* and *vitality*, both methods of analysis give the same results, and the results are very similar to the ones we obtain in the simulation study. The use of the BBR model, though it gives good results, leads to a loss of information in the categorization process of the dependent variable. Finally, HRQoL domains like *bodily pain*, *social functioning* and *mental health*, although reasonable continuous, are quite flat and do not fit well to either the normal or the beta-binomial distribution. One alternative to both methods in these continuous domains could be a non-parametric method such as bootstrap. However, Walters and Campbell [44] have recently concluded that bootstrap methods are not more appropriate for analysing HRQoL outcome data than standard methods based on normal distribution. In addition, bootstrap is not easily understood by the health care practitioners and it is not easy to perform with standard software.

Secondly, we compare both methods of analysis when the application focuses on the magnitude of the relationship between the covariates and HRQoL and its clinical interpretation and significance. Statistical significance itself does not provide concise information about clinically meaningful effects [45]. The process of defining clinical significance remains a challenge [46]. There are multiple strategies used to determine clinical significance among the various generic and disease specific functional health measures, including HRQoL questionnaires, and the issue is complex [47]. Ferguson *et al.* describe an attempt to develop a standard method of estimating clinically significant change when the SF-36 is used as a principal outcome and they present a psychometric method of estimating clinically significant SF-36 outcomes [48]. For instance, based on Ferguson *et al.*'s standards, the 27 points of difference in the *role-physical* domain between a patient with depression and a patient without depression detected by the MLR approach and shown in Table VIII is not a clinically significant difference. Twenty seven units is more or less one jump in the empirical ordinal scale for *role-physical* (each jump corresponds to 25 units). Thus, this difference is probably due to one unit difference in a single item, which is not enough to be considered clinically significant by Ferguson *et al.*'s standards. The same happens also with the 27 points of difference in the *role-emotional* domain due to anxiety. However, an odds ratio of four, based on coefficients shown in Table IX, is considered by experts in the field as clinically meaningful and highly informative. This basically means that a patient without depression is four times more likely to have a better *role-physical* than a patient with depression, or that a patient without anxiety is four times more likely to have a better *role-emotional* than a patient with anxiety. Therefore, both methods provide results that are interpreted differently and their clinical significance is evaluated in different ways. Whereas the MLR approach provides results that are not considered clinically significant in any of the eight HRQoL domains, based on Ferguson *et al.*'s standards; the BBR approach renders results interpretable in a way that is clinically significant, specially in *role-physical* and *role-emotional* domains.

Summarizing, for truly ordinal HRQoL domains, like *role-physical* and *role-emotional*, the BBR approach is more powerful than the MLR approach in detecting statistically significant covariates. Besides, the magnitude of the relationship between the covariates and these HRQoL domains and

DOI: 10.1002/sim

the interpretation of this relationship is quite different depending upon the selected method of analysis and different conclusions could be reached based on which one is considered clinically significant by researchers in the area. For the remaining HRQoL domains, the BBR approach detects the same significant covariates as the MLR and, although the magnitude of the relationship between the covariates and these HRQoL is quite differently interpreted, it does not affect clinical significance. In real applications of measuring HRQoL, a researcher is interested not only in detecting statistically significant relationships between HRQoL and other covariates, but also in the interpretation of the results and the clinical significance of such relationship. Considering that the eight domains of the SF-36 questionnaire must be analysed all together, the same method of analysis is preferable for all of them. In conclusion, the BBR approach is recommended to analyse HRQoL data measured by the SF-36 questionnaire, not only to detect statistically significant predictors of HRQoL, but also to measure the magnitude of such relationship and its clinical relevance.

As a limitation we need to mention the effect of categorization and the subjectivity of this process. Although categorization is not easy and natural for some of the continuous domains, specifically the *bodily pain* and *social functioning* domains, different approaches (from 9 to 12 categories) lead to very similar results. None of the observed differences between the MLR and the BBR results in the more continuous domains are important enough to prefer one over the other method. However, more research in the categorization process and sensitivity analysis is recommended and necessary.

As we have mentioned before, our work has concentrated primarily on observational studies where HRQoL has been measured as the main outcome and there are several confounding factors that might significantly affect HRQoL and, therefore, should be included in the model. It does not mean that this is not a valid methodology to analyse data from other settings such as clinical trials, but the performance of standard methods in this setting is broadly demonstrated [9, 12] and the benefit of using a more complicated method of analysis should be tested against simplicity in the particular context of clinical trials. As other authors have mentioned [49], the primary analysis of a major clinical trial should be relatively simple and familiar to a clinical audience, so the range of reasonable analysis is limited. If the goal of the analysis is to assess the magnitude of the treatment effect on the HRQoL outcome, then interest lies in comparing location between two treatments. In this particular setting, Heeren and D'Agostino [50] have demonstrated the robustness of the two-independent t-test applied to ordinal scales with less than five categories and this work has also been extended to account for a covariate in particular conditions [51, 52]. These results cannot be applied to our work, where the number of covariates that is necessary to incorporate in the models is more than one. In the particular case of the SF-36 outcome, modelling the SF-36 scores with the beta distribution, as proposed by Cheung *et al.* [29], could also be a good approach to compare location between two treatments. The objective of Cheung *et al.* is not the same we have here. We believe that more research along these lines is necessary, although that would be beyond the scope of the present paper.

In 1996, version 2.0 of the SF-36 was introduced [53]. It expanded from two to five the response choices for items in the two role functioning scales. Therefore, there are now more possible response categories for these two domains. Although the translation of the version 2.0 to Spanish has already been developed, it has not been validated yet [54].

Strictly speaking, our conclusions only apply to the SF-36 HRQoL outcome. Further work is required to test the better performance of BBR against standard methods for other HRQoL outcomes, populations or interventions.

DOI: 10.1002/sim

ACKNOWLEDGEMENTS

This work was partially supported by Universidad del País Vasco (UPV/EHU) (1/UPV/EHU 00100.310-EA-8108/2000 and UPV 00038.321-13631/2001), Departamento de Sanidad del Gobierno Vasco and Ministerio Español de Educación y Ciencia and FEDER (Research grant MTM2004-00341). The authors would like to thank Alvaro Muñoz for his helpful comments on an earlier research about the presented methodology. The authors thank two anonymous referees and the editor for providing thoughtful comments and suggestions, which led to substantial improvement of the presentation of the material in this paper.

REFERENCES

1. Testa MA, Simonson DC. Assessment of quality of life outcomes. *New England Journal of Medicine* 1996; **334**:835–840.
2. Goldsmith SB. The status of health status indicators. *Health Services Report* 1972; **87**:212–220.
3. Goldsmith SB. A reevaluation of health status indicators. *Health Services Report* 1973; **88**:937–941.
4. Katz S. The science of quality of life. *Journal of Chronic Diseases* 1987; **40**:459–465.
5. Spitzer WO. State of science 1986: quality of life and functional status as target variables for research. *Journal of Chronic Diseases* 1987; **40**:465–471.
6. Ware JE, Snow KK, Kosinski MA, Gandek B. *SF-36 Health Survey*, *Manual and Interpretation Guides*. The Health Institute, New England Medical Center: Boston, 1993.
7. Alonso J, Prieto L, Antó JM. La versión española del SF-36 health survey (Cuestionario de salud SF-36): un instrumento para la medida de los resultados clínicos. *Medicina Clinica* (*Barcelona*) 1995; **104**:771–776.
8. Schumacher M, Olschewski M, Schulgen G. Assessment of quality of life in clinical trials. *Statistics in Medicine* 1991; **10**:1915–1930.
9. Fairclough D. *Design and Analysis of Quality of Life Studies in Clinical Trials*. Chapman & Hall: London, 2002.
10. Olschewski M, Schumacher M. Statistical analysis of quality of life data in cancer clinical trials. *Statistics in Medicine* 1990; **9**:749–763.
11. Cox DR, Fitzpatrick R, Fletcher AE, Gore SM, Spiegelhalter DJ, Jones DR. Quality-of-life assessment: can we make it simple? *Journal of the Royal Statistical Society*, *Series A—General* 1992; **155**:353–375. (Disc: 375–393).
12. Fayers PM, Machin D. *Quality of Life*. *Assessment*, *Analysis and Interpretation*. Wiley: London, 2000.
13. Mesbah MF, Cole B, Ting Lee M. *Statistical Methods for Quality of Life Studies*. *Design*, *Measurements and Analysis*. Kluwer Academic Publishers: Boston, 2002.
14. Torrance GW. Utility approach to measuring health-related quality-of-life. *Journal of Chronic Diseases* 1987; **40**:593–600.
15. Lall R, Campbell MJ, Walters SJ, Morgan K, MRC CFAS. A review of ordinal regression models applied on health-related quality of life assessments. *Statistical Methods in Medical Research* 2002; **11**:49–67.
16. Walters SJ, Campbell MJ, Lall R. Design and analysis of trials with quality of life as an outcome: a practical guide. *Journal of Biopharmaceutical Statistics* 2001; **11**:155–176.
17. Van Hoeken D, Lucas AR, Hoek HW. Epidemiology. In *Neurobiology in the Treatment of Eating Disorders*, Hoek HW, Treasure JL, Katzman MA (eds). Wiley: London, 1998; 97–126.
18. Padierna A, Quintana JM, Arostegui I, González N, Horcajo MJ. The health-related quality of life in eating disorders. *Quality of Life Research* 2000; **9**:667–674.
19. Padierna A, Quintana JM, Arostegui I, González N, Horcajo MJ. Changes in health-related quality of life among patients treated for eating disorders. *Quality of Life Research* 2002; **11**:545–552.
20. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care* 1992; **30**:473–483.
21. Stansfeld SA, Roberts R, Foot SP. Assessing the validity of the SF-36 general health survey. *Quality of Life Research* 1997; **6**:217–224.
22. Brazier JE, Harper R, Jones NMB, O'Cathain A, Thomas KJ, Usherwood T, Westlake L. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *British Medical Journal* 1992; **305**:160–164.
23. Anderson JP, Kaplan RM, Coons SJ, Schneiderman LJ. Comparison of the quality of well being scale and the SF-36 results among two samples of ill adults: AIDS and other illnesses. *Journal of Clinical Epidemiology* 1998; **51**:755–762.
24. Wachtel T, Piette J, Mor V, Stein M, Fleishman J, Carpenter C. Quality of life in persons with human immunodeficiency virus infection: measurement by the medical outcomes study instrument. *Annals of Internal Medicine* 1992; **116**:129–137.

DOI: 10.1002/sim

25. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *Journal of Bone and Joint Surgery—American Volume* 1996; **78-B**:185–190.
26. Harper R, Brazier JE, Waterhouse JC, Walters SJ, Jones NMB, Howard P. Comparison of outcome measures for patients with chronic obstructive pulmonary disease (COPD) in an outpatient setting. *Thorax* 1997; **52**:879–887.
27. Rose MS, Koshman ML, Spreng S, Sheldom R. Statistical issues encountered in the comparison of health related quality of life in disease patients to published general population norms: problems and solutions. *Journal of Clinical Epidemiology* 1999; **52**:405–412.
28. Walters SJ, Campbell MJ. The use of bootstrap methods for analyzing health related quality of life outcomes (particularly the SF-36). *Health and Quality of Life Outcomes* 2004; **2**:70. DOI: 10.1186/1477-7525-2-70.
29. Cheung Y-B, Thumboo J, Machin D, Feng P-H, Boey M-L, Thio S-T, Fong K-Y. Modelling variability of quality of life scores: a study of questionnaire version and bilingualism. *Quality of Life Research* 2004; **13**:897–906.
30. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* (4th edn). American Psychiatric Association: Washington, DC, 1994.
31. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica* 1983; **67**:361–370.
32. Garner DM, Garfinkel PE. Eating attitudes test: an index of the symptoms of anorexia nervosa. *Psychological Medicine* 1979; **9**:273–279.
33. SAS Institute Inc. *SAS Procedures Guide*, *Version 6*. SAS Institute, Cary, NC, 1994.
34. MathSoft, Inc. *S-PLUS 2000 User's Guide*. MathSoft, Data Analysis Products Division, Seattle, WA, 1999.
35. Skellam JG. A probability distribution function derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *Journal of the Royal Statistical Society*, *Series B* 1948; **10**:257–261.
36. Chatfield C, Goodhart GJ. The beta-binomial model for consumer purchasing behaviour. *Applied Statistics* 1970; **19**:240–250.
37. Weil CS. Selection of the valid number of sampling units and a consideration of their combination in toxicological studies involving reproduction, teratogenesis and carcinogenesis. *Food and Cosmetics Toxicology* 1970; **8**:177–182.
38. Griffiths DAB. Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. *Biometrics* 1973; **29**:637–648.
39. Williams DA. The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics* 1975; **31**:949–952.
40. Crowder MJ. Beta-binomial ANOVA for proportions. *Applied Statistics* 1978; **27**:34–37.
41. Williams DA. Extra-binomial variation in logistic linear models. *Applied Statistics* 1982; **31**:144–148.
42. Morgan BJT. *Analysis of Quantal Response Data*. Chapman & Hall: London, 1992.
43. Cytel Software Corporation. *EGRET for Windows*. *User Manual*. Cytel Software Corporation, Cambridge, MA, 1999.
44. Walters SJ, Campbell MJ. The use of bootstrap methods for estimating sample size and analyzing health related quality of life outcomes. *Statistics in Medicine* 2005; **24**:1075–1102. DOI: 10.1002/sim.1984.
45. Kendall PC. Clinical significance. *Journal of Consulting and Clinical Psychology* 1999; **67**:283–284.
46. Clancy C, Eisenberg J. Outcomes research care: measuring the end results of health care. *Science* 1998; **282**:245–246.
47. Wyrwich K, Wollinsky F. Identifying meaningful intra-individual change standards for health-related quality of life measures. *Journal of Evaluation in Clinical Practice* 2000; **6**:39–49.
48. Ferguson RJ, Robinson AB, Splaine M. Use of the reliable change index to evaluate clinical significance in SF-36 outcomes. *Quality of Life Research* 2002; **11**:509–516.
49. White IR, Thompson SG. Choice of test comparing two groups, with particular application to skewed outcomes. *Statistics in Medicine* 2003; **22**:1205–1215. DOI: 10.1002/sim.1420.
50. Heeren T, D'Agostino RB. Robustness of the two independent samples *t*-test when applied to ordinal scaled data. *Statistics in Medicine* 1987; **6**:79–90.
51. Sullivan LM, D'Agostino RB. Robustness and power of analysis of covariance applied to data distorted from normality by floor effects: homogeneous regression slopes. *Statistics in Medicine* 1996; **15**:477–496.
52. Sullivan LM, D'Agostino RB. Robustness and power of analysis of covariance applied to ordinal scaled data as arising in randomized controlled trials. *Statistics in Medicine* 2003; **22**:1317–1334. DOI: 10.1002/sim.1433.
53. Ware JE, Kosinski M, Dewey JE. *How to Score Version 2.0 of the SF-36 Health Survey*. Quality Metric Incorporated, Lincoln, RI, 2000.
54. Vilagut G, Ferrer M, Rajmil L, Rebollo P, Permanyer-Miralda G, Quintana JM, Santed R, Valderas JM, Ribera A, Domingo-Salvany A, Alonso J. El Cuestionario de Salud SF-36 español: una década de experiencia y nuevos desarrollos. *Gaceta Sanitaria* 2005; **19**:135–150.