SORT 48 (2) July-December 2024, 155-182 DOI: 10.57645/20.8080.02.17

Patient-reported outcomes and survival analysis of chronic obstructive pulmonary disease patients: a two-stage joint modelling approach

Cristina Galán-Arcicollar[1,2], Josu Najera-Zuloaga[2] and Dae-Jin Lee[1,3]

Abstract

Joint modelling has gained attention in longitudinal studies incorporating biomarkers and survival data. In the context of chronic diseases, patient evolution is often tracked through multiple assessments, with patient-reported outcomes playing a crucial role. The Beta-Binomial distribution is suggested as a suitable model for these longitudinal variables. However, its integration into joint modelling remains unexplored. This study introduces an estimation procedure for analyzing longitudinal patient-reported outcomes and survival data together. We compare different estimation approaches through simulation experiments, including the proposed model. Furthermore, the methodologies are applied to real data from a follow-up study on chronic obstructive pulmonary disease patients.

MSC: *62N02, 62N03.*

Keywords: *Joint modelling, Beta-Binomial regression, Patient-reported outcomes, Survival analysis.*

1. Introduction

In recent years, there has been an increased focus on placing patients at the centre of health care research and evaluating clinical care (Weldring and Smith, 2013). For instance, patient-reported outcomes (PROs) are helpful tools that provide information on the patient's health status considering their health, quality of life, or functional status associated with the health care or treatment they received (Weldring and Smith, 2013). PROs have gradually become a significant source of information as they evaluate a wide

[1] Applied Statistics Research Line, Basque Center for Applied Mathematics, Bizkaia, Spain.
[2] Department of Mathematics, University of the Basque Country UPV/EHU, Bizkaia, Spain.
[3] School of Science and Technology, IE University, Madrid, Spain.

Received: June 2023
Accepted: December 2023

range of outcomes, such as pain, fatigue or vitality. Its use is strongly recommended in combination with clinical indicators to provide a more comprehensive patient assessment, especially in chronic illnesses (Speight and Barendse, 2010). This information is characterized by coming directly from the patient, without interpretation of the patient's response by a clinician or anyone else (US Department of Health and Human Services, 2006). The PRO measurements are usually collected by supplying validated questionnaires to patients. Thus, PROs are often built as a sum of responses to several items of the questionnaire, so they can be considered as discrete and bounded random variables. It has been shown in the literature that due to subject-specific characteristics, PROs are usually overdispersed (i.e., the mean-variance relationship fails mainly due to presence of an unexpected source of variation). In this context, the Beta-Binomial distribution has been proposed in the literature as an adequate distribution to fit overdispersed discrete and bounded outcomes, particularly in PRO analysis (Arostegui, Núñez-Antón and Quintana, 2007).

Most clinical studies that consider health-related quality of life (HRQoL) involve the follow-up of patients where longitudinal data are collected to evaluate patient worsening or changes in the health-status over time. Typically, it is also considered survival or time-to-event analysis during follow-up studies, when there is an event of interest such as death or illness relapse/recovery. The primary purpose of this article is to propose a methodology that allows answering the question of whether an association exists between the survival data and the serial measurements of HRQoL. In this framework, several studies have analyzed the impact of HRQoL on mortality (or survival) under a cross-sectional approach, as in Domingo-Salvany et al. (2002), or simply by fitting a model for the vital status instead for the survival times, such as Esteban et al. (2022). Other works have focused on the association of HRQoL with clinical measurements such as the number of hospitalizations or body mass index (BMI) in order to assess the evolution over time without considering survival data, like in Esteban et al. (2020). However, it is well known that when interest relies on the two outcomes (i.e., the longitudinal evolution and time-to-event), separate analyses are not the best modelling options because they do not consider the dependencies between them.

The statistical literature uses the term "joint modelling" to refer to those methods that simultaneously analyze longitudinal measurements and time-to-event outcomes. Nowadays, joint models of longitudinal and survival data have received much attention in the literature, particularly in medical studies, where these data frequently arise together in practice (Wu et al., 2012). The classical approach to joint modelling consists of a full likelihood formulation that integrates the two fitted models for longitudinal and survival to jointly estimate the parameter set. Regardless joint model's popularity, a complete analysis that includes longitudinal discrete and bounded outcome has not been thoroughly studied. Moreover, when these studies are carried out, the nature of data is not usually considered, where due to computational complexities, linearity or some kind of data transformation is assumed to simplify the computations, for example in Ibrahim, Chu and Chen (2010); Wu et al. (2012); Li, Tosteson and Bakitas (2013). In addition,

two-stage joint models consisting of estimating the longitudinal submodel and then plug the shared information into the survival submodel has been proposed as a simple solution in Self and Pawitan (1992). Although they are mainly known for their biased results when compared to fully likelihood methods (Wu et al., 2012), they are a well-known and a flexible methodology in the joint modelling framework. Bayesian methods have also been proposed to extend joint models to generalized linear mixed models, although it is desirable to check if the final results are sensitive to the choices of prior distributions (Armero, 2021).

In this work, although we do not overcome the possible bias of the two-stage methodology, we argue and show that it is preferable to consider a proper distribution for the longitudinal outcome rather than assuming a linear mixed model (i.e., a Gaussian response) when performing a joint model with a direct focus on PROs. Particularly, we aim to assess the effect of the longitudinal PRO measurements and the time until patient's death occurs. However, for the case of PROs, the fact that the Beta-Binomial distribution does not belong to the exponential family of distributions makes its inclusion into the joint modelling framework not straightforward. Our proposal provides an easy way to account for the evolution of PRO questionnaires of patients and the estimation of the survival probabilities by means of a joint modelling which includes the Beta-Binomial distribution for the analysis of longitudinal discrete and bounded data with overdispersion. A two-stage approach is considered in this work, where the first step consists of fitting a longitudinal Beta-Binomial mixed-effects model that estimates the impact of observed covariates and the subjects' evolution through time by following the approach described in Najera-Zuloaga, Lee and Arostegui (2019). Then, in a second step, the estimated linear predictors are included into a Cox proportional hazards regression model. We developed an unified methodology that couple these two models and show through simulation studies the better performance among other alternatives. Finally, in the suplementary material, we provide the R code to implement our approach using the functions `BBmm` in the R package `PROreg` (Najera-Zuloaga, Lee and Arostegui, 2022) for the longitudinal modelling of the PROs and `coxph` function of the `survival` R package (Therneau and Grambsch, 2000).

This research was motivated by an analysis of the health-status of patients with chronic obstructive pulmonary disease (COPD). Researchers at the Respiratory Service at Galdakao Hospital in Biscay (northern Spain) designed the COPD study, which was a longitudinal clinical trial that recorded measurements of the health status and evolution of patients being treated for COPD, see Esteban et al. (2020) for further details. One of the main objectives of the study was to measure the relationship between HRQoL and mortality of COPD patients. The hypothesis to assess this relationship is based on the fact that COPD is not only an airway obstruction disease, it is a complex, heterogeneous and multisystem disease (Vanfleteren et al., 2016) whose overall impact on individuals is many-sided. Thus, its severity is not fully captured by clinical parameters, it often needs to be supplemented by other indicators from a patient's perspective, such as those associated with HRQoL. This work is intended to provide clinicians and researchers on PROs

the statistical tools for modelling the HRQoL evolution of patients and its survival probabilities in an unified framework and discuss on joint modelling approaches in this context.

The paper is organized as follows. In Section 2, we introduce the details of our COPD study data set. Section 3 is dedicated to present the modelling of HRQoL evolution over time. In Section 4, we present two popular methodologies that incorporate longitudinal measurements in survival analysis and we provide a new approach for PRO data. The COPD data analysis with the presented methodologies is supplied in Section 5, with clinical interpretation of relevant results. In Section 6, we perform a simulation study based on the COPD data that compares the detailed methodologies in Section 4, including our proposal. Finally, in Section 7 it is given some conclusions and discussion.

2. Motivation study

Chronic obstructive pulmonary disease (COPD) is a major cause of chronic morbidity and mortality worldwide (Pauwels and Rabe, 2004). COPD is a respiratory system disease with irreversible damage which causes physiological discomfort and psychosocial impact on individuals, thus it is also associated with high level of disability. Although medical assessment of COPD mainly involves clinical biomarker measurements, the overall impact of COPD on individuals is multifaceted and it is not completely reflected by them. Thus, the assessment of the PROs and HRQoL is considered part of the standard care in the treatment of COPD (Wiklund, 2004). There is good evidence that COPD exacerbations can have a large and sustained impact on patients' symptoms and health status (Jones and Higenbottam, 2007). For this reason, tools such as PROs are needed to evaluate all different aspects of the disease, as they supplement clinical biomarkers by other kind of indicators from a patients perspective.

Our COPD study is an observational study that was designed at the Respiratory Unit at Galdakao Hospital in Biscay, Spain. A sample of 543 patients were consecutively included during the first year and in the second half of the study. The study is conducted in five years follow-up period for each patient with a maximum of four clinical examination and interviews per patient. Thus, the number of measurements per patient ranges between one and four. Figure 1 summarize this feature taking into account patients division according to the occurrence of event.

Notice that most patients had the maximum number of measurements but we also find that there are only one measurement recorded for some of them, being an unbalanced longitudinal data set. Moreover, patients' entry time, where baseline (first) measurement was recorded, did not take place at the same time due to consecutively entry of patients. In addition, measurement times were unequally space intervals because we found that the second measurement is one year apart from the fist one, as well as the third measurement from the second one, but the fourth measurement is three years apart from the third one.

The health-status in the COPD study was measured with both, generic and disease-specific questionnaires, named respectively Short Form-36 Health Survey (SF-36) and

St. George's Respiratory Questionnaire (SGRQ). Questionnaires often provide information about different health aspects, thus they are usually divided in dimensions according to the information referred. Particularly, SF-36 was constructed to represent eight health dimensions, which are the physical functioning (PF), role physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role emotional (RE), and mental health (MH). The first four dimensions are mainly physical, whereas the last four measure mental aspects of health. Each dimension scale score range from 0 to 100, where a higher score indicates a better health status. This standardized scoring system is detailed in Ware et al. (1993). On the other hand, SGRQ consist of three dimensions namely, symptoms (SYMP), impact (IMP) and activity (ACT) where higher scores refer to worse health status (Jones, Quirk and Baveystock, 1991). Each of the three dimensions of the questionnaire is separately standardized in the range 0 to 100.

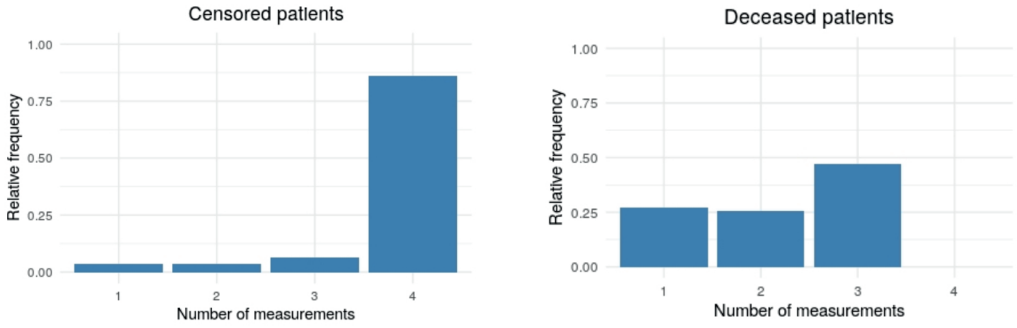

Figure 1. *Relative frequencies of patient's measurements. Left-side picture represents relative frequency of patients with no event while right-side represents relative frequency of patients with presence of the event during the study.*

In this work, we considered a re-scaled form of the patients' scores to interpret them in a binomial form, which makes its interpretation easier. The SF-36 re-scaling is motivated by the work of Arostegui, Núñez-Antón and Quintana (2013) and SGRQ is based on the idea that a 4-points change in the 0-100 scale is considered a clinically significant change (Jones, 2005). Based on this re-scaling, patients' average scores are shown in Table 1 according to the maximum score of each dimension.

In our COPD study, survival data was also collected and the time-to-event variable considered was the patient's date of death. Survival times are frequently influenced by right censoring in which the event time differs from the observed one. Particularly, in the COPD study administrative censoring was applied, which occurs when the study observational period ends without the presence of the event, i.e, when the patient has four measurements. Furthermore, we discovered censorship as a result of withdrawals, also known as loss to follow-up. During the study, 167 events were recorded, so 376 patients were censored. Because 324 patients completed the study's follow-up with no event, its censoring was administrative. Thus, 52 patients were censored as a result of the withdrawal process, corresponding approximately to a 10% of random censoring.

Table 1. *HRQoL scores of each test dimension are presented as mean (sd) for each year of the study. Number of patients at each year is also shown.*

		HRQoL mean measurements			
Test	**Dimension**	**Baseline**	**1 year**	**2 years**	**5 years**
	(max score)	n=543	n=484	n=425	n=320
SF-36	PF (20)	11.55 (4.88)	11.63 (4.99)	11.56 (4.94)	11.29 (4.99)
	RP (4)	2.62 (1.56)	2.44 (1.60)	2.51 (1.60)	2.22 (1.65)
	BP (9)	6.59 (2.58)	6.28 (2.71)	6.43 (2.66)	6.37 (2.57)
	GH (20)	8.81 (4.27)	8.54 (4.55)	8.32 (4.38)	8.22 (4.07)
	VT (20)	11.87 (4.99)	11.65 (4.80)	11.93 (4.68)	11.52 (4.78)
	SF (8)	6.53 (1.96)	6.39 (2.07)	6.57 (1.93)	6.23 (2.09)
	RE (3)	2.41 (1.08)	2.22 (1.18)	2.30 (1.14)	2.11 (1.23)
	MH (13)	9.47 (2.91)	9.42 (2.78)	9.44 (2.80)	9.25 (2.94)
SGRQ	SYMP (24)	10.63 (5.54)	10.11 (5.60)	10.35 (5.81)	10.52 (5.83)
	IMP (24)	7.52 (5.21)	7.11 (5.29)	7.06 (5.08)	7.12 (5.21)
	ACT (24)	11.63 (6.19)	10.95 (6.19)	11.21 (6.14)	11.34 (6.31)

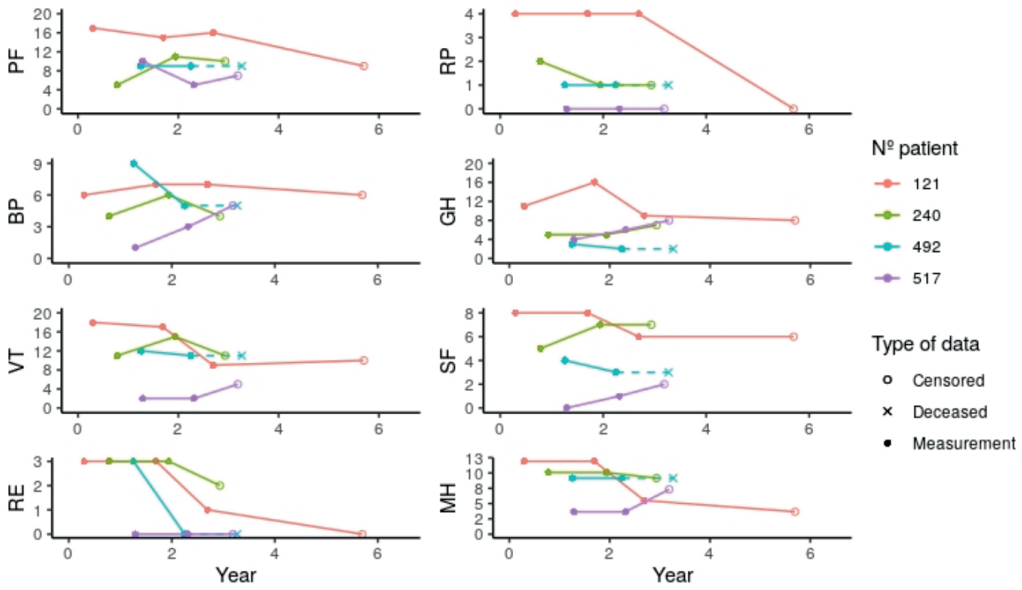

Figure 2. *SF-36 longitudinal measurements according to the survival data recorded of four COPD patients. The straight dashed line indicates the extrapolation of the patient's score from the last recorded measurement to the observed date of death.*

Figure 2 and Figure 3 illustrate the longitudinal measurements of the eight SF-36 dimensions and the three SGRQ dimensions respectively of four patients of the database.

According to the survival data recorded, we indicated the dates of death with a cross symbol and joined with the previous measurements with a dashed line because at patient's date of death no HRQoL data was recorded. This is an example of the different types of data recorded according to the two kinds of censorship and the event recorded. For instance, patients 240 and 517 only recorded 3 measurements and then no more information was collected, so they are censored due to loss of follow-up, while patient 121 completed the four measurements and is administratively censored.

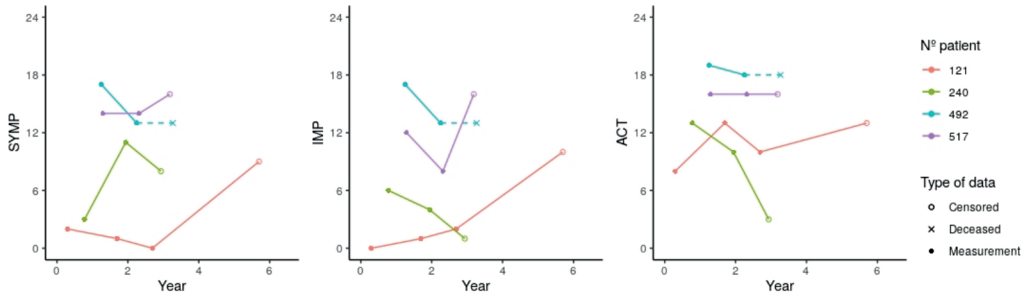

Figure 3. *SGRQ longitudinal measurements according to the survival data recorded of four COPD patients. We used the straight dashed line to joint the date of death with the recorded measurements.*

3. Modelling the evolution of patients' HRQoL

This article is focused on self-reported outcomes, which usually have discrete and bounded distributions. A particular case of self-reported outcomes in medicine are the patient-reported outcomes (PROs). It is known that this kind of measurements usually lead to floor or ceiling effect, i.e, typically accumulate values in one or both edges of the score scale due to subject-specific characteristics, leading to different shapes such as U, J or inverse J-shaped (Najera-Zuloaga, Lee and Arostegui, 2018). The usual exponential family distributions are not able to fit them properly and particularly, the normality assumption for the longitudinal variable in classic joint models will not be frequently satisfied for this kind longitudinal data. The Beta-Binomial distribution has been proposed in literature as the proper distribution to analyze discrete and bounded outcomes with overdispersion (Arostegui et al., 2007). Generally, the Beta-Binomial distribution is defined as a mixture of the Binomial and Beta distributions. It consists of a finite sum of Bernoulli variables whose probability parameter is random and follows a Beta distribution. The Beta-Binomial distribution preserves the characteristics of the Binomial distribution which suits the nature of discrete and bounded data, but it also displays the flexibility of the Beta distribution. We denote that variable Y follows a Beta-Binomial distribution as $Y \sim BB(m, p, \phi)$ with parameters m, p and ϕ, where parameter m makes reference to the maximum number of trials, p is the probability parameter and ϕ the correlation/dispersion parameter. The density function of the Beta-Binomial distribution does not belong to the exponential family distributions but it has a closed-form equation,

see Arostegui et al. (2007) for further details. The article Najera-Zuloaga et al. (2018) proposes a marginal regression model with the Beta-Binomial distribution for measuring the effect of explanatory variables on discrete and bounded response variables. This Beta-Binomial regression is performed by connecting the probability parameter to a vector of regression parameters by means of a logit link function model.

PROs are usually measured in a longitudinal framework in which individuals are followed up for a certain period. The extension of Beta-Binomial regression models to the longitudinal framework is performed in terms of mixed-effects that associate all measurements for the same subject. Beta-Binomial mixed-effect model (BBMM) includes random effects into the linear predictor to accommodate the dependency of repeated measurements. As in mixed model methodology, conditioned on the random effects, the repeated measurements are independent and drawn from a Beta-Binomial distribution. Let $y_i = (y_{i1}, ..., y_{in_i})$ be the repeated measurements for subject i, then $y_{ij}|u_i \sim BB(m_i, p_{ij}, \phi)$, $\forall j \in \{1, ..., n_i\}$, and u_i its corresponding vector of random effects $u_i \sim N(0, D)$ that describe the subject-specific characteristics. The probability parameter of the Beta-Binomial distribution is also linked with logit link function to the fixed and random parameters such as:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = x_{ij}\beta + z_{ij}u_i = \eta_{ij}, \quad i = 1, ..., n_i, \tag{1}$$

where x_{ij} and z_{ij} are the fixed and random covariates respectively for the jth measurement of subject i with β and u_i the corresponding effect parameters. See Najera-Zuloaga et al. (2019) for further details.

4. Relating longitudinal patient reported outcomes and survival analysis

During follow-up studies in clinical trials, it is of particular interest to collect several biomarker measurements as well as time-to-event outcomes, such as death, illness relapse, recovery or development of some disease (Papageorgiou et al., 2019). Thus, longitudinal data and survival data frequently appear together in practice, and they are often associated in some ways (Wu et al., 2012). The longitudinal measurement of biomarkers is useful for characterizing the occurrence of an outcome of interest because they can predict treatment outcomes or be related to the event process and prognosis (Arisido et al., 2019). It is then of particular interest to evaluate and investigate its relationship (Ibrahim et al., 2010). In this section, we present two well-known methodologies for estimating the relationship between these outcomes, as well as define our proposed approach.

4.1. A Time-Varying Cox-Model

The first proposed approach to analyze the relationship between longitudinal and survival outcomes consisted of extending the classical proportional hazards model (Ri-

zopoulos, 2012). The main objective was to allow the inclusion of time-dependent covariates into this known survival model. This methodology is usually referred as Time-Varying Cox-Model (TVCM). The risk function for patient i is defined quite similar to the original Cox model and it can be written as follow:

$$h_i(t|\mathcal{Y}_i(t), w_i) = h_0(t)R_i(t)\exp\{\gamma w_i + \alpha y_i(t)\}, \tag{2}$$

where $h_i(t)$ and $h_0(t)$ respectively denote the subject and baseline risk function at each time t, $\mathcal{Y}_i(t) = \{y_i(s) : s \leq t\}$ is the history of longitudinal covariate up to time t and w_i refers to the vector of baseline covariates for patient i. Parameters γ and α measure the impact of baseline and longitudinal covariates respectively into the risk function. Finally, it is included a risk indicator function $R_i(\cdot)$ with $R_i(t) = 1$ if subject i is at risk at time t, and $R_i(t) = 0$ otherwise.

TVCM is known for being a flexible semi-parametric methodology for fitting survival analysis because parameters estimation is based on proportional hazards. Thus, partial likelihood is used to perform parameters estimation, where baseline risk function is left unspecified. However, it is assumed in TVCM that time-dependent covariates are predictable processes, measured without error, and have their complete path fully specified (Rizopoulos, 2012). Notice that the whole longitudinal history is not available because information is recorded only at some measurement times. In order to overcome this issue, TVCM is based on the so-called 'last value carried forward' (LVCF) approach, where the marker values are considered constant between measurement points. In Figure 4, we considered the PF and MH dimensions of SF-36 COPD data shown in Figure 2 and we showed a graphic interpretation of the stepwise approach of the TVCM methodology. We can observe that the longitudinal trends are not considered in this method.

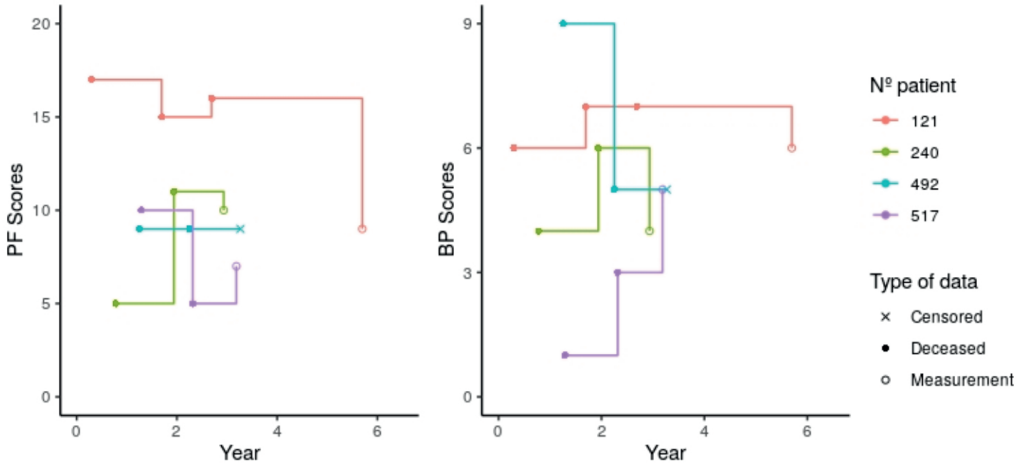

Figure 4. *Graphic interpretation of the last value carried forward used in Time-Varying Cox Model.*

Moreover, it is important to distinguish between the different kinds of time-varying covariates because it is known that TVCM is not appropriate when time-dependent covariates are of endogenous nature (Rizopoulos, 2012). Endogenous covariates, such as biomarkers, typically require the survival of the subject for their existence and thus, they satisfy $S_i(t|\mathcal{Y}_i(t)) = Pr(T_i^* > t|\mathcal{Y}_i(t)) = 1$, where $S_i(\cdot)$ and T_i^* make reference to the survival function and the random time-to-event variable for subject i respectively. Then, for instance, if the date of death for the subject is s, the covariate will not be recorded at each time $t \geq s$. This fact of endogenous covariates is not taken into account in TVCM. In addition, despite TVCM flexibility, several studies have performed a variety of simulated scenarios to show that it can lead to really biased results due to all assumptions mentioned above, see for instance Arisido et al. (2019).

The implementation of this methodology in R software makes used of displayed data sets using the counting process format $(start, stop, status)$, holding the information on the longitudinal variable for a specific time interval. The start and stop times denote the limits of the time intervals during which measurements were recorded and the status indicate the ocurrence of the event at the end of the interval. The `tmerge` function aids in the creation of such data sets and the model is fitted in the `coxph` function, both functions from the `survival` package (Therneau and Grambsch, 2000).

4.2. Shared parameter joint modelling

In order to overcome TVCM limitations, an alternative modelling framework has been introduced in the literature, known as the joint modelling for longitudinal and time-to-event data (Faucett and Thomas, 1996). The motivating idea behind these joint models is to couple the survival model with a suitable model for the repeated endogenous measurements. The longitudinal model is usually fitted by using Linear Mixed Model (LMM) to describe covariates impact together with time evolution, and survival model is usually performed in terms of the Cox model, leading to a definition of risk function as follows:

$$h_i(t|\mathcal{M}_i(t), w_i) = h_0(t)\exp\{\gamma w_i + \alpha m_i(t)\}. \tag{3}$$

It reminds quite similar to the subject's risk function defined in TVCM shown in Equation (2), but the values included in the survival model are not the observed ones $y_i(t)$, instead it is incorporated the term $m_i(t)$ that denotes the true and unobserved value. This true value is computed according to the longitudinal model assumed, thus:

$$m_i(t) = \mathbb{E}(y_i(t)|u_i) = x_i(t)\beta + z_i(t)u_i, \tag{4}$$

where $x_i(t)$ are the fixed parameters with β the corresponding fixed effect and $z_i(t)$ the random parameters with u_i the corresponding vector of random effects having a multivariate normal distribution with mean zero and covariance matrix D, i.e, $u_i \sim N(0, D)$. One of the most popular estimation method that has been proposed for the described joint model consisted of computing a full maximum likelihood of the observed outcomes. Both models, longitudinal and survival, are set and then, they are linked using a shared

latent structure, due to this feature joint models belong to the class of shared-parameter models. The key assumption of this methodology is full conditional independence, i.e, it is assumed that conditional to random effects, time-to-event and longitudinal outcomes are independent, as well as the different measurements for the same individual (Rizopoulos, 2012). Considering the classical survival analysis notation, let T_i be the recorded time-to-event and δ_i the event indicator, that takes value 1 if the observed time T_i equals to the event time T_i^* and 0 otherwise, the assumptions can be written as follow:

$$f_\theta(T_i, \delta_i, y_i|u_i) = f_\theta(T_i, \delta_i|u_i) f_\theta(y_i|u_i),$$

$$f_\theta(y_i|u_i) = \prod_{j=1}^{n_i} f_\theta(y_i(t_{ij})|u_i),$$

where θ denotes the set of all, survival and longitudinal parameters, i.e., $\theta = (\theta_t, \theta_y, \theta_u)$ and y_i denotes the vector for the n_i measurements for subject i taken at time points $\{t_{ij} : j = 1, ..., n_i\}$. Under these assumptions the log-likelihood contribution for the ith subject can be formulated in this way:

$$\begin{aligned}\log f_\theta(T_i, \delta_i, y_i) &= \log \int f_\theta(T_i, \delta_i, y_i|u_i) f_{\theta_u}(u_i) du_i \\ &= \log \int f_{\theta_t,\beta}(T_i, \delta_i|u_i) \big[\prod_j f_{\theta_y}(y_i(t_{ij})|u_i)\big] f_{\theta_u}(u_i) du_i.\end{aligned}$$

As it was mentioned, time-to-event outcome is usually fitted by Cox model and the density is included in the log-likelihood characterized by the hazard and survival functions:

$$\begin{aligned}f_{\theta_t}(T_i, \delta_i|u_i) &= h_i(T_i|\mathcal{M}_i(T_i))^{\delta_i} S_i(T_i|\mathcal{M}_i(T_i)) \\ &= \big[h_0(T_i)\exp\{\gamma w_i + \alpha m_i(T_i)\}\big]^{\delta_i} \times \exp\left(-\int_0^{T_i} h_0(s)\exp\{\gamma w_i + \alpha m_i(s)\} ds\right).\end{aligned}$$

Notice that for joint modelling, it is necessary to set a baseline hazard function. Weibull or exponential functions are frequently used as baseline hazard in survival analysis, but also it can be modeled flexibly using piecewise-constant or B-splines (Rosenberg, 1995) methods. Furthermore, it is necessary to set the longitudinal model to compute $m_i(\cdot)$ according to Equation (4) and also for the inclusion of the density function into the full likelihood, which is the normal density function. The literature also considered the use of generalized linear mixed models (GLMM) for longitudinal data into the joint modelling framework. This extension is straightforward (Rizopoulos, 2012) by choosing the density function as a member of the exponential family. The difficulty of this extension of joint models to incorporate GLMM is that computation becomes more demanding because of the nonlinearity of the longitudinal models (Wu et al., 2012). Moreover, this extension is not implemented in the available R software packages. Due to that fact, it is common to assume linearity measurements, sometimes ignoring the nature of data, being Joint Modelling with longitudinal normality assumption one of the

widely used methodology in the literature. Moreover, it is worth mentioning that the inclusion of several longitudinal responses into the model also leads to computational complexities, being this another limitation.

The implementation of this methodology can be easily performed by using `jointModel` function from `JM` R-package (Rizopoulos, 2010), it is needed first to fit separately the linear mixed-effects model with `lme` from `nlme` package (Pinheiro et al., 2013) and Cox model with `coxph` from `survival` package (Therneau and Grambsch, 2000). In this article, we denote this methodology as JM.

4.3. A two-stage approach for discrete bounded outcomes

In this subsection, we present our proposal approach where we incorporate the Beta-Binomial distribution into the joint modelling framework by performing a two-stage methodology. Our main goal is to emphasize the need of an adequate model that assesses the relationship between longitudinal and a time-to-event outcomes which provides an easy interpretation for PROs. With that aim we propose a joint model methodology that includes Beta-Binomial distribution for a suitable fit of this kind of longitudinal data, which relies on a better estimation of the association parameter between the longitudinal and the time-to-event outcome.

The first step consists of fitting a longitudinal Beta-Binomial mixed-effects model to evaluate the impact of some observed covariates and also its evolution over time including subject-specific effects that account for non observable characteristics that are different for each individual. Then, in a second step, the estimated linear predictor is computed according to the Beta-Binomial regression model and it is included in a Cox proportional hazards survival model as observed covariates.

For a sample of N individuals, let $y_i = (y_{i1}, ...y_{in_i})$ be the n_i measurements for subject i such that $y_{ij}|u_i \sim BB(m, p_{ij}, \phi)$, taken at measurement times $t_i = (t_{i1}, ...t_{in_i})\ \forall i = 1, ..., N$. Then, following BBMM approach the probability parameter and the linear predictor are linked by means of *logit* function as indicated in Equation (1), where p_{ij} denotes the probability parameter for patient i at time t_{ij}, i.e., $p_{ij} = p_i(t_{ij})$. Fixed covariates x_{ij}, can be time-dependent or baseline covariates and random effects $u_i \sim \mathcal{N}(0, D(\sigma))$ are assumed to follow a multivariate Normal with zero mean and variance-covariance nonsingular matrix D, which depends on a vector of variance parameters σ.

In this first step where BBMM is fitted, parameters β, u_i, σ and ϕ are estimated. The analysis of these parameters provides an assessment of the impact of fixed effects, heterogeneity among subjects and overdispersion of the longitudinal responses. Next, for the second step, the parameter estimations are used to compute the fitted values for the longitudinal outcome at each time t following the Beta-Binomial methodology. Let p_{it} denote the probability parameter of subject i at each time t and $\hat{\eta}_{it}$ its corresponding estimated linear predictor, then:

$$\hat{y}_i(t) = m\hat{p}_{it} = m \cdot \text{logit}^{-1}(\hat{\eta}_{it}) = \frac{m}{1 + \exp(-\hat{\eta}_{it})}. \tag{5}$$

Once the fitted values are computed, in the second step, they are inserted into the classic Cox model as if they were observed values. Thus, we proceed by fitting the Cox model as usual survival analysis defining the risk function as:

$$\lambda_i(t) = \lambda_0(t)\exp(\gamma w_i + \alpha \hat{y}_i(t)). \tag{6}$$

In this study, we focused on the main idea of estimating joint models using two-stage approaches, which avoid integral computation difficulties and also allow a greater flexibility when setting the longitudinal model. Additionally, this methodology avoids the transformation of the HRQoL questionnaire scores by means of square root, logarithmic or Box-Cox. These data transformation of the questionnaries' scores is usually performed to set linearity assumption, although this approach makes interpreting coefficients more difficult. The inclusion of BBMM allows the analysis of the association parameter with hazard ratio interpretation in terms of 1-point scoring, which is quite intuitive and natural instead of incorporating the standardized data or data transformations. Moreover, the performance of a BBMM longitudinal regression provides an easy way to interpret covariates' influence for 1-unit increment in terms of the odds-ratio, akin to hazard ratio analysis.

Our proposal can be easily implemented with the available R software. The first step is performed by using `BBmm` function from `PROreg` package (Najera-Zuloaga et al., 2022) and second step is fitted with `coxph` function from `survival` package (Therneau and Grambsch, 2000). See the R code provided in the supplementary material.

5. Application to COPD study

In the COPD study, one of the main objectives was to describe HRQoL and its evolution. This objective was studied in Esteban et al. (2020), where they analyzed the impact of sociodemographic variables and clinical indicators on the HRQoL, although only the SF-36 questionnaire was considered. Another key objective was to collect survival data based on time-to-event and the corresponding event indicators to assess HRQoL's relation to patient's risk of death. This objective was considered for SGRQ scores in Esteban et al. (2022). However, in this work they did not perform a survival analysis, because a logistic regression was fitted for vital status and only one-year periods were considered without including the time-to-event nature of the study. The original study aimed to investigate both, how HRQoL evolved and its relationship with the survival data collected. Nevertheless, these two outcomes have not been thoroughly studied jointly, where the inclusion of BBMM into the joint modelling framework may improve the results of the association parameter and provide an easier clinical interpretation. It is our goal to provide a complete analysis of both outcomes of interest for a more comprehensive evaluation. Apart from a joint analysis of both outcomes being preferable to a separate one, we also want to emphasize the importance of PRO data being adequately fitted.

The first step of our proposal can deal with one of the main objectives of the COPD study, which consists of measuring the evolution of patients' HRQoL scores over time

to determine trends. We performed a Beta-Binomial mixed-effect model by considering time as observable covariate to evaluate population and subject-specific evolution by including random intercept and slope as follow:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = (\beta_0 + u_{i0}) + (\beta_1 + u_{i1})\, t_{ij} = \eta_{ij}, \tag{7}$$

such that the vector of random effects satisfies $u_i = (u_{i0}, u_{i1}) \sim \mathcal{N}(0, D(\sigma))\ \forall i = 1, ..., 543$, with D a diagonal matrix with entries $\sigma = (\sigma_{u_0}, \sigma_{u_1})$. The components of vector u_i represent respectively baseline and subject-specific effects in the evolution of patient i. Results are shown in Table 2 for the fixed slope and the standard deviation of random effects. It should be mentioned that the algorithm failed to estimate the variance of the random slopes, σ_{u_1}, for the dimensions BP, SF, MH and IMP. In fact, for the aforementioned dimensions, due to the limited number of measurements per subject, the model is not able to capture differences in the longitudinal trends of individuals. Therefore, we adjusted the model by removing the random slopes effects from the linear predictor and evaluated it again in these dimensions.

Table 2. *Univariate BBMM results. Fixed slope parameter estimates are shown, each with standard error, p-value and odds-ratio (OR) according to one-unit increment in time covariate. Standard deviations of random effects (intercept and slope) are also included with the corresponding standard errors. The symbol * represents tendency to zero deviance.*

		Fixed coefficients			Random sd	
		β_1 (se)	p-value	OR	σ_{u_0} (se)	σ_{u_1} (se)
SF36	PF	-0.06 (0.01)	<0.001	0.94	1.03 (0.03)	0.12 (0.01)
	RP	-0.13 (0.03)	<0.001	0.88	1.67 (0.06)	0.31 (0.02)
	BP	-0.03 (0.02)	0.1347	0.97	1.16 (0.04)	*
	GH	-0.04 (0.01)	<0.001	0.96	0.80 (0.03)	0.06 (0.00)
	VT	-0.04 (0.01)	<0.001	0.96	0.92 (0.03)	0.10 (0.01)
	SF	-0.07 (0.02)	<0.001	0.93	1.53 (0.06)	*
	RE	-0.10 (0.04)	0.007	0.90	2.19 (0.10)	0.53 (0.03)
	MH	-0.03 (0.01)	0.004	0.97	0.99 (0.03)	*
SGRQ	SYMP	0.01 (0.01)	0.499	1.01	0.76 (0.03)	0.10 (0.01)
	IMP	0.00 (0.01)	0.670	1.00	1.06 (0.03)	*
	ACT	0.03 (0.01)	0.001	1.03	1.18 (0.04)	0.13 (0.01)

The results of both questionnaires, generic (SF-36) and disease-specific (SGRQ), show that the health status of the patients is worsening over time. For instance, one year of evolution in the RP dimension is associated with an odds-ratio of 0.88 (i.e. $\exp(-0.13) \approx 0.88$). As a result, for each year of evolution, the patient is approximately 12% less likely to score one more point in the RP dimension, which means that RP patients' scores will decrease resulting in poorer health status.

Table 3. *Univariate Cox model results for hazard ratio including 95% confidence intervals. JM included standardized 0-100 data divided in order to compare results with TSBB and TVCM included binomial form data. Significant results according to p-value are in bold.*

SF-36 scores association with patients' mortality						
	TSBB		TVCM		JM	
	HR	95% CI	HR	95% CI	HR	95% CI
PF (20)	**0.91**	**0.88, 0.94**	**0.88**	**0.84, 0.90**	**0.87**	**0.83, 0.90**
RP (4)	0.96	0.85, 1.09	**0.82**	**0.74, 0.90**	**0.74**	**0.63, 0.88**
BP (9)	0.97	0.88, 1.07	**0.90**	**0.85, 0.96**	0.91	0.81, 1.02
GH (20)	0.98	0.93, 1.02	**0.94**	**0.90, 0.98**	**0.94**	**0.90, 0.99**
VT (20)	**0.94**	**0.90, 0.98**	**0.92**	**0.89, 0.95**	**0.91**	**0.87, 0.95**
SF (8)	0.94	0.84, 1.03	**0.86**	**0.80, 0.93**	**0.82**	**0.72, 0.93**
RE (3)	0.93	0.80, 1.08	**0.79**	**0.70, 0.90**	**0.69**	**0.54, 0.87**
MH (13)	0.96	0.90, 1.03	**0.89**	**0.85, 0.94**	**0.92**	**0.85, 0.99**

Once longitudinal parameters are estimated, fitted values are computed and included in a classic Cox model to evaluate the relationship between HRQoL scores evolution and patient's risk of death. Therefore, the second step of our proposal deals with the second objective of the COPD study, which consisted of evaluating the association between HRQoL results and patient's risk of death. We considered both questionnaires in the survival analysis. First, for SGRQ, we considered a multivariate approach that incorporates the questionnaire's three dimensions in the survival model. However, due to the existing correlation between the dimensions, parameter interpretation can be misleading as it is shown in Table 4 and will be explained in the next paragraphs. It is worth mentioning that the Cox model can easily handle multiple longitudinal covariates, but `JM` package can only incorporate one longitudinal covariate per model and, thus, only univariate models can be performed. Second, in order to avoid misleading interpretation due to covariate correlation and as an illustrative way to show the diversity in the covariate inclusion that the proposed model offers, for generic SF-36 questionnaire we considered an univariate survival approach and thus, we incorporated each dimension separately into the survival models. Furthermore, we included the transformed binomial form data in TVCM for its easy interpretation, but we considered the standardized 0–100 data in the JM methodology because it is more intuitive to assume a normality-based longitudinal model. However, to facilitate a comparison of the estimations of the different methods, we divided the standardized 0–100 scores such that each dimension's maximum score coincides with that in its binomial form.

Results are displayed in Table 3 for the SF-36 survey and Table 4 for the SGRQ. The results are shown according to the association coefficient α of Equation (6), significant results are in bold and they are graphically shown in a forest plot in Figure 5 and 6 for SF-36 and SGRQ respectively. The association coefficients are interpreted in terms of Hazard ratio (HR$=\exp(\alpha)$). Thus, at any particular time point t, HR denotes the

relative increase in the risk at time t that results from one unit increase in the longitudinal variable at the same time point. This interpretation is akin to the odds-ratio interpretation of coefficients in Beta-Binomial regression.

Table 4. *Multivariate Cox model results for hazard ratio including 95% confidence intervals. JM included standardized 0–100 data divided in order to compare results with TSBB and TVCM that included binomial form data. Significant results according to p-value are in bold. (*) Univariate models were considered in JM.*

SGRQ scores association with patients' mortality						
	TSBB		TVCM		JM (*)	
	HR	95% CI	HR	95% CI	HR	95% CI
SYMP (24)	**0.94**	**0.89, 0.99**	0.98	0.94, 1.01	1.03	0.99, 1.07
IMP (24)	0.99	0.93, 1.06	0.98	0.93, 1.04	**1.07**	**1.03, 1.10**
ACT (24)	**1.11**	**1.06, 1.17**	**1.11**	**1.06, 1.17**	**1.09**	**1.06, 1.13**

The results show that the PF (measures mobility disability) and VT (measures energy and fatigue) dimensions were statistically significant in predicting patients' risk of death according to our TSBB methodology. Because PF and VT are two dimensions of the generic SF-36 test, higher scores are associated with lower risk of death, with PF having the greatest impact. In particular, our TSBB proposal relates one more point in the PF dimension with a 9% lower risk. These two significant dimensions of the SF-36 deal with daily patients' activity and how do they feel about it, so the patients' perception about their physical activity is related with a significant impact on mortality.

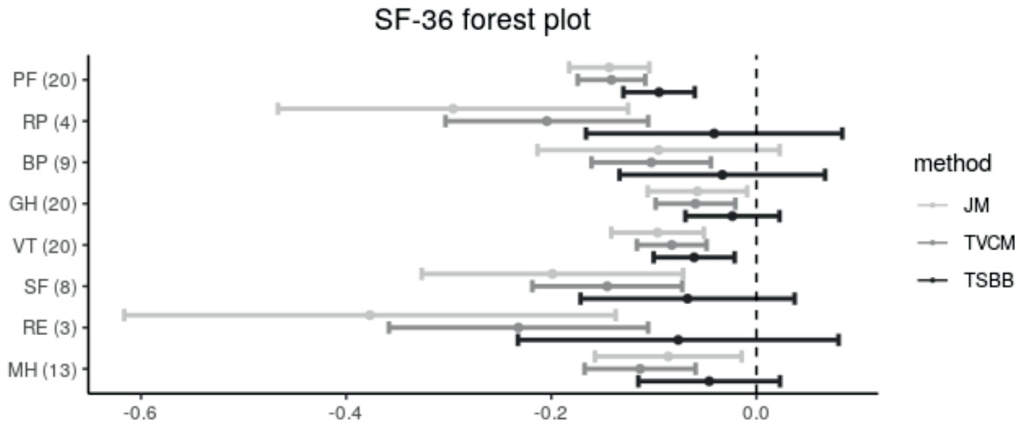

Figure 5. *Forest plot of association coefficient for each SF-36 dimension according to each methodology.*

TVCM results are quite different because all dimensions had signficative effect on patients' risk. It is important to remark that, as it was mentioned in Section 4.1, this methodology extends the previous longitudinal outcome until the event occurrence and

therefore less variability is assumed. This leads into lower standard deviation and smaller confidence intervals, which can be the source because of all the results remain significant.

JM methodology is dealing with standardized data that we transformed in order to compare the one-point change in the dimensions' score and thus, results are mainly different in all dimensions that those obtained from TVCM and TSBB. The results of JM for PF and VT dimensions showed high decrease in patient's risk for one more point scored in these dimensions compared to TSBB and TVCM. This fact will be due to overestimation of this methodology shown in the simulation study in Section 6.

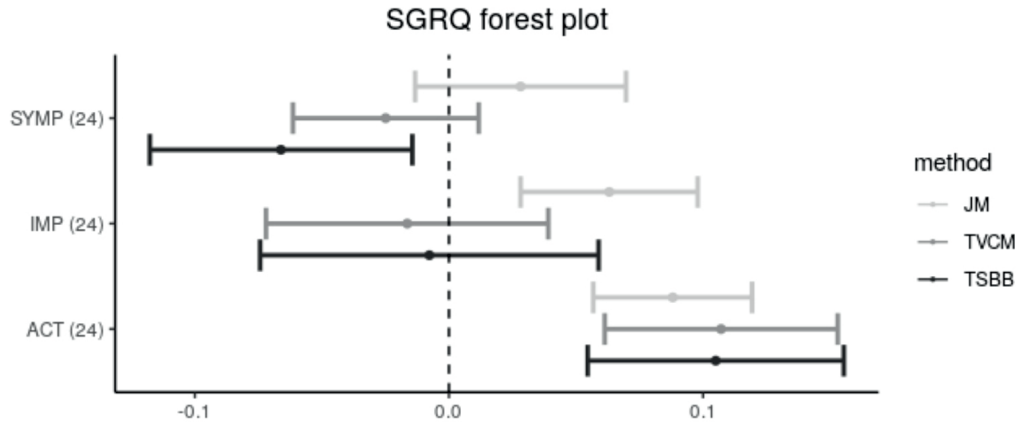

Figure 6. *Forest plot of association coefficient for each SGRQ dimension according to each methodology.*

The SGRQ scores are related with lower health status of the subjects. Particularly, the ACT dimension (the effect of disturbances to mobility and physical activity) was significant in the three methodologies having the greatest impact on patients mortality. Our TSBB methodology and TVCM coincides in associating one more point in the score of this dimension with 11% higher risk, being a 9% risk increment in JM method. We found that the TSBB multivariate model associates an additional point in the SYMP dimension (quantifying distress due to respiratory symptoms) with 6% lower risk. This misleading effect occurs because of covariate correlation. In fact, both approaches that incorporate the three dimensions in the same model, TSBB and TVCM, attribute a decreasing risk effect to the increment of SYMP dimension, although in TCVM the effect is not statistically significant. On the other hand, when JM univariate analysis was performed, IMP dimension showed significant effect on risk where one added point is associated with 7% higher risk of death. Finally, we can conclude that in both questionnaires, the patients' perception of their physical activity and how they feel about it could be a great indicator to take into account in COPD patients' mortality.

6. Simulation Study

In this section, we present a simulation study to assess the performance of the proposed method and compare it with two methodologies widely used in the literature to deal with the analysis of the parameter association between the longitudinal and time-to-event outcomes: TVCM and JM with longitudinal normality assumption. Generally, in literature, TVCM is performed when the main interest is to assess the relationship between the two outcomes because of its popularity and flexibility, leaving the longitudinal model unspecified. The performance of the joint model based in normality assumption is the most commonly used in literature even for discrete and bounded longitudinal data, where standardization or data transformation is performed, avoiding data nature. The simulations have two main aims: (a) validate our proposal approach under several parameter conditions and (b) compare the performance of these three approaches in controlled scenarios offering a variability of situations according to longitudinal shape and the relationship among outcomes.

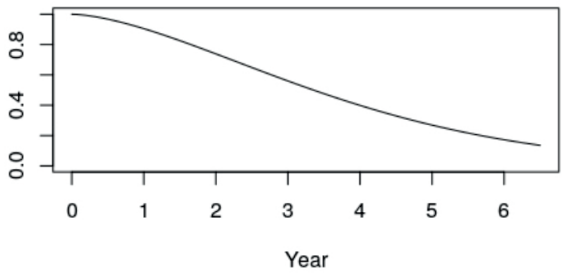

Figure 7. *Baseline survival function for the Weibull baseline hazard with scale and shape parameters* $w_1 = 0.1$, $w_2 = 1.6$

The overall scenario settings are mainly based on COPD study, which was detailed in Section 4. We are considering the same maximum number of measurements per patient, which is four. The longitudinal outcome was computed at the entry and measurement times, dispersed as in the COPD study. Thus, the measurements times are generated as follow: The first measurement corresponds to the subject's entry time which is performed according to a uniform distribution in the interval $(0, 1.5)$. The second measurement is one year apart from the first, as well as the third measurement from the second. However, the fourth measurement is three years apart from the third measurement. The described times are not equally spaced, as in the COPD study, where the data are collected at irregularly spaced times. The overall follow-up period is of five years since the simulated subject's entry.

Time-to-event or survival times, denoted as T_i^*, are generated by evaluating the inverse of a cumulative hazard from Equation (3), see Crowther and Lambert (2013). To that aim, we assumed the Cox proportional hazard model with Weibull baseline risk function $h_0(t) = w_1 w_2 t^{w_2 - 1}$, where w_1 and w_2 denote the scale and shape parameter re-

spectively. See Figure 7 for the parameters $w_1 = 0.1$ and $w_2 = 1.6$. By choosing this baseline hazard function we aim to simulate a baseline survival function that don't reach 0 at the end of the follow-up period, as not all patients die at the end of the COPD study.

We also considered an administrative censoring time A_i according to the COPD study, such that if they reach the fourth measurement, the five-year follow-up period ends without event observation, thus it satisfies $A_i = t_{i4}$. Besides, in order to perform possible dropouts, a loss of follow-up censoring time C_i is performed with a uniform distribution between patient time entry and patient last measurement time with around 10% of censored individuals, which is the censoring rate we found in the COPD study. Finally, the observed time for each subject is computed as $T_i = \min\{T_i^*, A_i, C_i\}$. Furthermore, an event indicator δ_i is recorded, i.e, if $T_i = T_i^*$, then $\delta_i = 1$ and 0 otherwise. Generated marker values $y_i(t_{ij})$ with $t_{ij} > T_i$ were disregarded.

The experiment consisted of 200 random simulations with 250 subjects. We considered a maximum of four measurements per subject, resulting in 1000 observations per simulation of a longitudinal variable distributed as a Beta-Binomial with a fixed maximum score m, a probability parameter p, and a dispersion parameter ϕ. For the sake of clarity, probability parameter p is computed according to the model assumed in Equation (7), considering subjects' overall and specific evolution over time.

Fitted values in Equation (3) of the survival model are computed according to the Beta-Binomial distribution model following Equation (5), which using linear predictor from Equation (7) leads to:

$$m_i(t) = \frac{m}{1 + \exp\{-((\hat{\beta}_0 + \hat{u}_{i0}) + (\hat{\beta}_1 + \hat{u}_{i1})t)\}}. \tag{8}$$

In order to achieve both objectives of the simulation study, all longitudinal parameters were set except for the dispersion parameter ϕ and the association parameter α of the survival model. By giving different values to the dispersion parameter ϕ, longitudinal data can adopt a wide variety of shapes and varying parameter α allows us to evaluate a small, moderate and strong association between the two outcomes of interest.

Two main scenarios were considered in order to simulate longitudinal responses based on two dimensions of the HRQoL questionnaires considered in the COPD study, one dimension of the SF-36 and another one of the SGRQ. This scenario division will allow us to evaluate positive and negative association among outcomes, like it happens in the SF-36 and the SGRQ respectively. In one scenario we considered 24 as response maximum score and positive association parameter while in the other 8 was set up as maximum score and negative association parameter. See Table 5 for a summary of both main scenarios.

Table 5. *Setting parameters to perform two main simulation scenarios.*

	β_0	β_1	σ_{u_0}	σ_{u_1}	m	α
Scenario 1	-0.19	0.03	1.2	0.05	24	> 0
Scenario 2	0.40	-0.15	1.5	0.3	8	< 0

Once the two main scenarios were set, we provided scenario division by varying dispersion parameter using $\phi \in \{0.05, 0.5, 1\}$, so that, longitudinal distribution takes different shapes. See Figure 8 and 9 for a graphic image of the different generated shapes for Scenario 1 and Scenario 2 respectively. Notice that we have generated bell, flat, U, J and inverse J shapes.

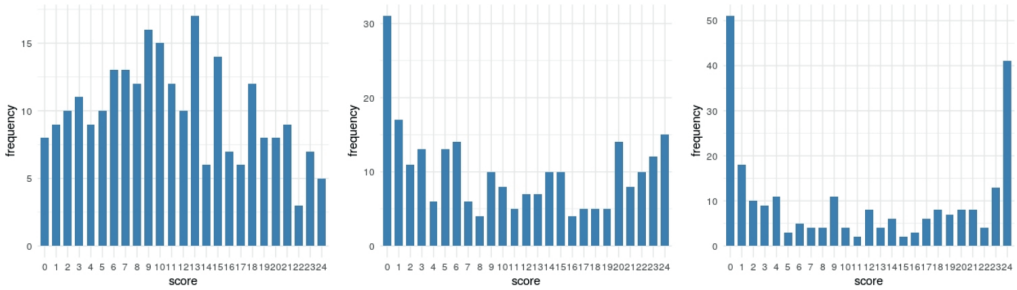

Figure 8. *Frequency of the simulated longitudinal scores based on Scenario 1 with* $\phi = 0.05, 0.50, 1$ *from left to right.*

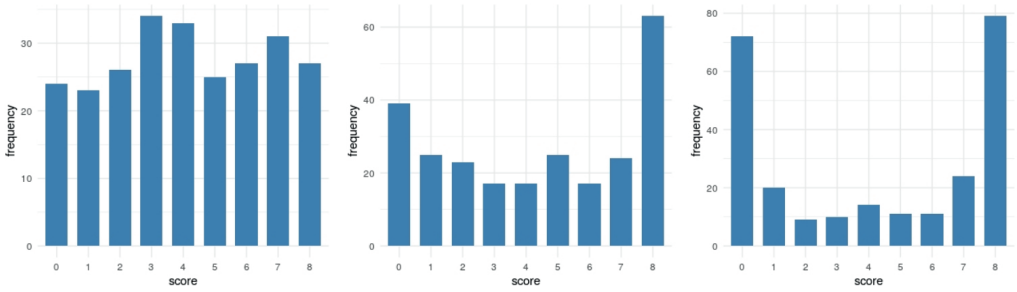

Figure 9. *Frequency of the simulated longitudinal scores based on Scenario 2 with* $\phi = 0.05, 0.50, 1$ *from left to right.*

Furthermore, we want to compare the performance of the methods by varying the relationship between the two outcomes of interest and thus the association parameter was set in order to perform scenarios corresponding to small, moderate and strong association between $m_i(t)$ and $h_i(t)$. As association parameter interpretation is done considering the hazard ratio in terms of scoring one additional point, it is important to take into account the maximum score number, because increasing an additional point does not have the same effect when the maximum score is 8 or 24. Thus, according to maximum number of trials, we set the association parameter as $\alpha \in (0.01, 0.05, 0.10)$ for Scenario 1 and $\alpha \in (-0.05, -0.10, -0.15)$ for Scenario 2. These modifications of ϕ and α lead to 9 sub-scenarios for each of the two main set above. Methods are represented as TSBB (Two-Stage Beta-Binomial) for our proposed methodology, TVCM for Time-Varying Cox Model and JM for Joint Model normality-based.

Although the estimates of all parameters (longitudinal and survival) have been obtained by both joint model methodologies, only results for α will be shown in detail.

First, because the association parameter between outcomes is the one estimated by the three methodologies, as TVCM does not provide a longitudinal analysis. Besides, longitudinal coefficients cannot be entirely compared for JM and TSBB because the longitudinal model differs. Lastly, because the simulation objective is to compare the methods' performance when estimating the relationship between the two outcomes of interest. Moreover, related to baseline hazard function, as TSBB and TVCM do not need to specify it, we set piecewise method in JM baseline hazard selection in order to allow a flexible semi-parametric methodology.

The results of the simulation study are summarised in Table 6 for Scenario 1 and Table 7 for Scenario 2. Results in both tables are based on the following statistics: the *%Bias* ($bias/\alpha$), the empirical standard deviation (ESD), the average standard deviation (ASD) and the 95% coverage probability (CP). It is worth mentioning that convergence problems related to the normality assumption in the longitudinal model were found when applying the JM methodology. Mainly it occurred in Scenario 1 for $\alpha = 0.1$, where there is a higher risk of death and, therefore, fewer longitudinal data. The convergence percentage was only 34% for $\phi = 0.5$ sub-scenario, which makes this methodology hardly applicable in case studies. Only in Scenario 2 for $\alpha = -0.05$ and $\phi = 0.05$ this methodology converged 100% times. To compare the three different approaches, we contemplated those 200 realizations where all methods converged.

Table 6. *Results of the association parameter α obtained from the proposed method (TSBB), the TVCM and joint model (JM) fitted to data considering Scenario 1 in Table 5 with $\alpha \in (0.01, 0.05, 0.10)$ and $\phi \in (0.05, 0.5, 1)$. Percentage bias (%Bias), empirical standard deviation (ESD), average standard deviation (ASD) and 95% coverage probability (CP) are shown.*

		Scenario 1											
		$\alpha = 0.01$				$\alpha = 0.05$				$\alpha = 0.10$			
ϕ		*%Bias*	ESD	ASD	CP	*%Bias*	ESD	ASD	CP	*%Bias*	ESD	ASD	CP
0.05	TSBB	0.11	0.012	0.012	98.0	0.37	0.014	0.012	72.0	0.36	0.018	0.014	33.0
	TVCM	0.18	0.010	0.011	96.0	0.12	0.012	0.011	89.0	0.05	0.011	0.012	94.0
	JM	0.43	0.012	0.013	95.5	0.48	0.016	0.014	63.5	0.45	0.018	0.018	33.5
0.5	TSBB	-0.01	0.013	0.013	96.0	0.23	0.018	0.013	76.0	0.09	0.025	0.014	73.5
	TVCM	-0.37	0.008	0.008	92.5	-0.36	0.009	0.008	38.5	-0.47	0.008	0.008	0.0
	JM	0.46	0.015	0.016	95.0	0.51	0.022	0.019	75.0	0.24	0.034	0.024	84.0
1	TSBB	-0.16	0.015	0.012	86.5	0.04	0.016	0.012	85.0	-0.10	0.025	0.012	60.0
	TVCM	-0.47	0.008	0.007	88.5	-0.50	0.008	0.007	7.5	-0.59	0.007	0.007	0.0
	JM	0.50	0.018	0.017	92.5	0.48	0.024	0.021	87.5	0.21	0.046	0.029	77.5

First, both tables show that TVCM has the lowest ESD in all scenarios due to the LVCF approach that assumes longitudinal outcomes constant among measurement times. According to Table 6, we can see that this methodology produces highly biased

results, specially when the dispersion and association parameters are increased, as evidenced by the poor CP results, which reached zero in some sub-scenarios. Table 7 demonstrates that TVCM produced less biased results when there is small association between $m_i(t)$ and $h_i(t)$. Otherwise, its results are generally not the least biased in the moderate and strong association scenarios, which, in addition to its low ESD, is reflected in its lower CP when compared to other methods.

For the TSBB approach, the ESD remains quite similar in all the scenarios of Table 6, being slightly larger when dispersion or association parameters increase, which also happens in Table 7. This conclusion results quite logical as increasing variability in the outcomes also increases the uncertainty of the estimations. Its ESD results are higher than the ones in TVCM in all sub-scenarios because fitting longitudinal outcomes when the event takes place includes more variability rather than considering the last value recorded constant. Concerning bias, TSBB presents the lowest *%Bias* in Table 6 for most of the scenarios, and it does so in Table 7 except for the small-association case, where only around 33 results were statistically significant. However, in Table 6, we find low CP compared with other methodologies, mainly in sub-scenario $\alpha = 0.1$ that might be due to high patient risk, that few longitudinal data is taken into account and coefficient estimation is poorer. Then, there is greater variability in estimates of the longitudinal sub-model. However, in the second step we include the fitted value in the Cox model without incorporating the estimated variability of the longitudinal part. Then the difference between the ESD and the ASD produces the low CP values in those sub-scenarios.

Table 7. *Results of the association parameter α obtained from the proposed method (TSBB), the TVCM and joint model (JM) fitted to data considering Scenario 2 in Table 5 with $\alpha \in (-0.05, -0.10, -0.15)$ and $\phi \in (0.05, 0.5, 1)$. Percentage bias (%Bias), empirical standard deviation (ESD), average standard deviation (ASD) and 95% coverage probability (CP) are shown.*

		Scenario 2											
		$\alpha = -0.05$				$\alpha = -0.10$				$\alpha = -0.15$			
ϕ		*%Bias*	ESD	ASD	CP	*%Bias*	ESD	ASD	CP	*%Bias*	ESD	ASD	CP
0.05	TSBB	-0.32	0.033	0.031	90.5	0.09	0.032	0.033	95.0	0.23	0.037	0.038	89.0
	TVCM	0.17	0.031	0.028	90.0	0.26	0.032	0.031	87.5	0.30	0.035	0.034	75.0
	JM	0.74	0.037	0.035	81.5	0.78	0.039	0.039	51.5	0.83	0.049	0.046	23.0
0.5	TSBB	-0.41	0.027	0.030	91.5	-0.02	0.031	0.033	97.5	0.07	0.035	0.037	96.0
	TVCM	-0.12	0.023	0.024	95.0	-0.09	0.026	0.026	95.0	-0.08	0.029	0.029	92.5
	JM	0.79	0.038	0.040	85.0	0.80	0.045	0.045	59.5	0.85	0.056	0.055	32.0
1	TSBB	-0.43	0.029	0.029	88.0	-0.12	0.032	0.032	94.0	0.00	0.035	0.035	95.0
	TVCM	-0.23	0.022	0.022	91.0	-0.23	0.023	0.024	86.0	-0.19	0.026	0.027	81.0
	JM	0.87	0.044	0.043	84.5	0.83	0.053	0.050	62.5	0.92	0.069	0.061	33.5

Lastly, the JM approach ends with the highest *%Bias* in almost all the scenarios, specially in Table 7 where the maximum score number of the longitudinal outcome was set equal to 8. Normality assumption tends to fail as the maximum score number is lower in binomial data. Therefore, the simulation study shows the importance of considering an adequate distribution for the longitudinal outcome. Moreover, we can see that JM results are overestimated in all sub-scenarios which, in fact, the same effect can be observed in Section 5 where real data estimations with JM methodology were more extreme compared to those obtained with TSBB or TVCM methodologies.

Based on the simulation study results, we can conclude that although our proposal is based on a two-stage methodology, which is known as a biased approach, it performed better estimations in most scenarios. The first reason is that TVCM results are more biased in most cases and tend to underestimate the association parameter. Lastly, a Joint Model with a normality-based assumption for longitudinal data leads to really skewed results and high ESD. Moreover, it has significant convergence problems due to the normality assumption of the discrete and bounded longitudinal data. Hence, we emphasize the importance of considering a complete analysis of both outcomes and remark on the consequences of considering an inaccurate distribution of longitudinal data.

7. Conclusion and further research

In biological or clinical trials, which usually involve the follow-up of subjects, longitudinal and time-to-event variables are often recorded. Traditionally, although interest relied primarily on clinical indicators, PRO measurements have gained relevance in recent years and are now highly recommended in patients' assessments (Deshpande et al., 2011). This tool allows researchers to collect patient-perceived information about several topics, usually regarding health-related quality of life (HRQoL). Several studies are arising to examine HRQoL as a surrogate indicator of prognosis concerning mortality, showing the usefulness of HRQoL in the prediction of prognosis (Esteban et al., 2022). However, we found that some analysis were performed under a cross-sectional framework that did not take into account the repeated examinations and others that are only focused on measure patient's HRQoL evolution without considering survival data.

In the literature, we realized that there are two common methodologies for analyzing the relationship between longitudinal and time-to-event outcomes: Time-Varying Cox Model (TVCM) and the classic Joint Model based on shared-random effects (JM). TVCM incorporates the repeated measurements as covariates considering its value constant between the measurement time intervals, whereas JM jointly fits both outcomes where it is considered a longitudinal normality-based model. However, these two popular approaches are not able to incorporate PRO characteristics. Firstly because TVCM is not considered adequate for endogenous measurements that require the survivality of the patient to be recorded. Lastly, because the classic Joint Model leads to computational complexities if normal distribution is not assumed for the longitudinal data. To fill this gap in the state of the art, the main objective of this work is to perform a complete anal-

ysis of longitudinal HRQoL measurements and survival data that incorporates the PRO discrete, bounded and overdispersed essence.

In this study, we proposed a joint model based on two-stage methodology that allows the inclusion of the Beta-Binomial distribution for longitudinal data and also avoids computational complexities. We focused on the idea of a suitable fitting for repeated measurements that enhance the estimation of the relationship between longitudinal and survival data. Moreover, we emphasized how critical it is to select a wrong longitudinal model when considering a joint model performance. Our proposal includes a Beta-Binomial mixed-effect model to accommodate the correlation structure of longitudinal overdispersed discrete and bounded outcomes. In addition, the semi-parametric Cox model is set for survival data where longitudinal fitted values are included as covariates. Although the two stage approach is not a novel methodology for the joint modelling framework, the estimation procedure that we propose is the first one in the literature that entirely incorporates the nature of the PRO data.

This approach also provides a longitudinal fit of the PRO data that TVCM does not consider and also set an adequate model for repeated PRO measurements that causes computational complexities in the JM methodology. Moreover, several longitudinal PRO measurements can be easily incorporated in the survival model which is not possible in the JM approach. To compare the performance of our proposed methodology with those that are widely used in the literature for assessing the relationship between longitudinal PRO measurements and survival data, we carried out a simulation study. The simulation study showed that in most of the scenarios, the proposed method obtained better performance with a smaller bias and a greater coverage probability of the parameter than the other methods. Besides, results showed that JM presented significant convergence problems due to normality assumption, which makes it hardly applicable for real data. This method also lead to really biased estimations. The TVCM was mainly affected by PRO overdispersion because when variability is increased in the longitudinal outcomes its bias worsens due to the last value carried forward assumption.

We also applied the methodologies to the detailed COPD study to analyze real data and show that our methodology presents an easy interpretation of the results in terms of odds-ratio and hazard ratio. Relevant conclusions were exposed for both generic and disease-specific questionnaires. It was shown that the patient's perception of its activity levels and its feelings about it has a significant impact on the patient's risk of death.

Our proposed methodology has potential limitations. The two-stage methodology is mainly known for being a biased approach and it is shown in the simulation study results seemed to be biased. Moreover, a few number of repeated measurements per subject can lead into a to poor estimates of the longitudinal model that will be reflected in the estimation of the association parameter. Furthermore, the second step of the process does not account for the variability of longitudinal estimations, which can underestimate the variance of the estimations as it was shown in simulation study by ESD and ASD differences. Consequently, it sometimes results in low CP results despite of having a low bias. For further work, we aim to investigate a Bayesian approach that allow us to incorpo-

rate Beta-Binomial distribution for repeated measurements and jointly fits longitudinal and survival models. We also aim to include the longitudinal estimation variability in order to avoid low CP results and correct the potential bias that standard two-stage methods present. In the literature, several proposals exist for mechanisms to reduce the bias inherent in the two-stage joint modeling methodology while preserving its main advantages: the ability to address complex structures flexibly and the reduced computational demand. See, for instance, the works of Leiva-Yamaguchi and Alvares (2021) and Alvares and Leiva-Yamaguchi (2023).

As concluding remarks, we recommend the used of our proposed two-stage methodology that incorporates a Beta-Binomial mixed-effects model into the joint modelling framework. This methodology provides a complete analysis of longitudinal PRO and survival data with an adequate fit over popular methodologies.

Acknowledgements

This research is supported by AEI and EJ-GV under Grants PID2020-115882RB-I00 / AEI / 10.13039/5011 00011033 with acronym "S3M1P4R", Departamento de Educacion, Política Lingüística y Cultura del Gobierno Vasco [IT1456-22] and also by the Basque Government through the BERC 2022-2025 program and by the Ministry of Science and Innovation: BCAM Severo Ochoa accreditation CEX2021-001142-S / MICIN / AEI / 10.13039/501100011033. We also gratefully acknowledge Dr. Cristobal Esteban for providing the COPD data and Inmaculada Arostegui for the scientific discussion.

References

Alvares, D. and Leiva-Yamaguchi, V. (2023). A two-stage approach for Bayesian joint models: reducing complexity while maintaining accuracy. *Statistics and Computing*, 33(5):1–11. ISSN: 15731375. DOI: 10.1007/s11222-023-10281-9. URL: https://doi.org/10.1007/s11222-023-10281-9.

Arisido, M. W., Antolini, L., Bernasconi, D. P., Valsecchi, M. G. and Rebora, P. (2019). Joint model robustness compared with the time-varying covariate Cox model to evaluate the association between a longitudinal marker and a time-to-event endpoint. *BMC Medical Research Methodology*, 19(1):1–13. ISSN: 14712288. DOI: 10.1186/s12874-019-0873-y.

Armero, C. (2021). Bayesian Joint Models for Longitudinal and Survival Data. *Wiley StatsRef: Statistics Reference Online*, pages 1–6. DOI: https://doi.org/10.1002/9781118445112.stat08129.

Arostegui, I., Núñez-Antón, V. and Quintana, J. M. (2007). Analysis of the short form-36 (SF-36): The beta-binomial distribution approach. *Statistics in Medicine*, 26(June):1318–1342. DOI: 10.1002/sim.

Arostegui, I., Núñez-Antón, V. and Quintana, J. M. (2013). On the recoding of continuous and bounded indexes to a binomial form: An application to quality-of-

life scores. *Journal of Applied Statistics*, 40(3):563–582. ISSN: 02664763. DOI: 10.1080/02664763.2012.749845.

Crowther, M. J. and Lambert, P. C. (2013). Simulating biologically plausible complex survival data. *Statistics in Medicine*, 32(23):4118–4134. ISSN: 0277-6715. DOI: 10.1002/sim.0000.

Deshpande, P. R., Sudeepthi, B. L., Rajan, S. and Nazir, C. P. A. (2011). Patient-reported outcomes: A new era in clinical research. *Perspectives in Clinical Research*, 2(4):137. ISSN: 2229-3485. DOI: 10.4103/2229-3485.86879.

Domingo-Salvany, A., Lamarca, R., Ferrer, M., Garcia-Aymerich, J., Alonso, J., Félez, M., Khalaf, A., Marrades, R. M., Monsó, E., Serra-Batlles, J. and Antó, J. M. (2002). Health-related quality of life and mortality in male patients with chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, 166(5):680–685. ISSN: 1073449X. DOI: 10.1164/rccm.2112043.

Esteban, C., Arostegui, I., Aramburu, A., Moraza, J., Aburto, M., Aizpiri, S., Chasco, L. and Quintana, J. M. (2022). Changes in health-related quality of life as a marker in the prognosis in COPD patients. *ERJ Open Research*, 8(1):00181–2021. DOI: 10.1183/23120541.00181-2021. URL: http://dx.doi.org/10.1183/23120541.00181-2021.

Esteban, C., Arostegui, I., Aramburu, A., Moraza, J., Najera-Zuloaga, J., Aburto, M., Aizpiri, S., Chasco, L. and Quintana, J. M. (2020). Predictive factors over time of health-related quality of life in COPD patients. *Respiratory Research*, 21(1):1–11. ISSN: 1465993X. DOI: 10.1186/s12931-020-01395-z.

Faucett, C. L. and Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine*, 15(15):1663–1685. ISSN: 02776715. DOI: 10.1002 / (SICI) 1097-0258(19960815)15:15¡1663::AID-SIM294¿3.0.CO;2-1.

Ibrahim, J. G., Chu, H. and Chen, L. M. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16):2796–2801. ISSN: 0732183X. DOI: 10.1200/JCO.2009.25.0654.

Jones, P. and Higenbottam, T. (2007). Quantifying of severity of exacerbations in chronic obstructive pulmonary disease adaptations to the definition to allow quantification. *Proceedings of the American Thoracic Society*, 4(8):597–601. ISSN: 15463222. DOI: 10.1513/pats.200707-115TH.

Jones, P. W. (2005). St. George's respiratory questionnaire: MCID. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 2(1):75–79. ISSN: 15412555. DOI: 10.1081/COPD-200050513.

Jones, P. W., Quirk, F. H. and Baveystock, C. M. (1991). The St George's Respiratory Questionnaire. *Respiratory Medicine*, 85:25–31. ISSN: 09546111. DOI: 10.1016/S0954-6111(06)80166-6.

Leiva-Yamaguchi, V. and Alvares, D. (2021). A two-stage approach for bayesian joint models of longitudinal and survival data: Correcting bias with informative prior. *Entropy*, 23(1):1–10. ISSN: 10994300. DOI: 10.3390/e23010050.

Li, Z., Tosteson, T. D. and Bakitas, M. A. (2013). Joint modeling quality of life and survival using a terminal decline model in palliative care studies. *Statistics in Medicine*, 32(8):1394–1406. ISSN: 02776715. DOI: 10.1002/sim.5635.

Najera-Zuloaga, J., Lee, D.-J. and Arostegui, I. (2018). Comparison of beta-binomial regression model approaches to analyze health-related quality of life data. *Statistical Methods in Medical Research*, 27(10):2989–3009. ISSN: 14770334. DOI: 10.1177/0962280217690413.

Najera-Zuloaga, J., Lee, D.-J. and Arostegui, I. (2019). A beta-binomial mixed-effects model approach for analysing longitudinal discrete and bounded outcomes. *Biometrical Journal*, 61(3):600–615. ISSN: 15214036. DOI: 10. 1002 / bimj . 201700251.

Najera-Zuloaga, J., Lee, D.-J. and Arostegui, I. (2022). *PROreg: Patient Reported Out- comes Regression Analysis*. R package version 1.2. URL: https://CRAN.R-project.org/package=PROreg.

Papageorgiou, G., Mauff, K., Tomer, A. and Rizopoulos, D. (2019). An overview of joint modeling of time-to-event and longitudinal outcomes. *Annual Review of Statistics and Its Application*, 6(August 2018):223–240. ISSN: 2326831X. DOI: 10.1146/annurev-statistics-030718-105048.

Pauwels, R. A. and Rabe, K. F. (2004). Burden and clinical features of chronic obstructive pulmonary disease (COPD). *Lancet*, 364(9434):613–620. ISSN: 01406736. DOI: 10.1016/S0140-6736(04)16855-4.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and R Core Team (2013). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-108.

Rizopoulos, D. (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1–33. ISSN: 15487660. DOI: 10.18637/jss.v035.i09.

Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data With Applications in R*. CRC Press, p. 274. ISBN: 1439872864.

Rosenberg, P. S. (1995). Hazard Function Estimation Using B-Splines. *Biometrics*, 51(3):874. ISSN: 0006341X. DOI: 10.2307/2532989.

Self, S. and Pawitan, Y. (1992). Modeling a Marker of Disease Progression and Onset of Disease. *AIDS Epidemiology*, (1991):231–255. DOI: 10.1007/978-1-4757-1229-2_11.

Speight, J. and Barendse, S. M. (2010). FDA guidance on patient reported outcomes. *BMJ (Online)*, 341(7772):518. ISSN: 17561833. DOI: 10.1136/bmj.c2921.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer, New York. ISBN: 0-387-98784-3.

US Department of Health and Human Services (2006). Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims: Draft guidance. *Health and Quality of Life Outcomes*, 4:1–20. ISSN: 14777525. DOI: 10.1186/1477-7525-4-79.

Vanfleteren, L. E. G. W., Spruit, M. A., Wouters, E. F. M. and Franssen, F. M. E. (2016). Management of chronic obstructive pulmonary disease beyond the lungs.

The Lancet Respiratory Medicine, 4(11):911–924. ISSN: 22132619. DOI: 10.1016/ S2213-2600(16)00097-7. URL: http://dx.doi.org/10.1016/ S2213-2600(16)00097-7.

Ware, J. E., Snow, K. K., Kosinski, M. and Gandek, B. (1993). SF36 Health Survey, Manual and Interpretation Guides.

Weldring, T. and Smith, S. M. S. (2013). Article Commentary: Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Services Insights*, 6:61–68. ISSN: 11786329. DOI: 10.4137/HSI.S11093.

Wiklund, I. (2004). Assessment of patient-reported outcomes in clinical trials: The example of health-related quality of life. *Fundamental and Clinical Pharmacology*, 18(3):351–363. ISSN: 07673981. DOI: 10.1111/j.1472-8206.2004.00234.x.

Wu, L., Liu, W., Yi, G. Y. and Huang, Y. (2012). Analysis of longitudinal and survival data: Joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 2012. ISSN: 1687952X. DOI: 10.1155/2012/640153.