Joint Modelling of Complex Longitudinal Outcomes and Time-to-Event Data

With a Focus on Quality of Life Outcomes in Cancer Clinical Trials

Hortense Doms

Thesis submitted in partial fulfillment of the requirements for the Degree of Doctor in Sciences

April 2026

Institute of Statistics, Biostatistics and Actuarial Sciences (ISBA)
Louvain Institute of Data Analysis and Modeling in Economics and Statistics (LIDAM)
UCLouvain, Louvain-la-Neuve, Belgium

Thesis Committee:

Pr. **Philippe Lambert** (Supervisor)	ULiège & UCLouvain, Belgium
Pr. **Catherine Legrand** (Supervisor)	UCLouvain, Belgium
Pr. **Anouar El Ghouch** (President)	UCLouvain, Belgium
Pr. **Karim Barigou** (Secretary)	UCLouvain, Belgium
Dr. **Murielle Mauer**	EORTC, Belgium
Dr. **Virginie Rondeau**	INSERM, France

Acknowledgments

Je souhaite tout d'abord remercier chaleureusement mes promoteurs de thèse, Catherine Legrand et Philippe Lambert, pour m'avoir donné l'opportunité de poursuivre un doctorat, mais surtout pour m'avoir accompagnée afin de mener à bien cette thèse. Merci pour le temps que vous y avez consacré, pour vos conseils et vos enseignements, mais surtout pour la confiance que vous m'avez accordée tout au long de ces années.

Je tiens également à remercier les membres de mon jury de thèse : Anouar El Ghouch, Karim Barigou, Murielle Mauer et Virginie Rondeau. Je vous remercie pour le temps que vous avez consacré à la lecture de ce manuscrit ainsi que pour vos remarques constructives et vos conseils enrichissants, qui ont contribué à améliorer la qualité de ce travail.

Durant ces années, la vie à l'institut n'aurait pas été la même sans mes collègues, anciens et actuels. Je remercie l'ensemble des chercheurs et assistants, ainsi que les membres de l'ISBA et du SMCS, pour avoir contribué à une ambiance de travail aussi agréable et bienveillante. Un grand merci tout particulier à mes collègues et amis avec qui j'ai partagé le bureau D.373 : Benjamin, Hugo, Aigerim et Ensiyeh. Vous avez été un précieux soutien au quotidien. Nos discussions et les moments partagés ont été une véritable parenthèse durant cette thèse. Vous m'avez aidée à rester motivée et positive, même dans les périodes plus difficiles. Je remercie également les professeurs avec qui j'ai eu l'opportunité d'encadrer des cours pratiques pour leur confiance.

Je n'oublie pas ma famille, dont le soutien a été essentiel. Plus spécialement, je remercie mes parents, mes beaux-parents ainsi que mes sœurs Ophélie, Lauraine et Éléonore. Merci pour votre bienveillance, vos encouragements et vos mots réconfortants. Tous ensemble et chacun à votre manière, vous avez largement contribué à l'aboutissement et à la réussite de cette thèse.

Finalement, le plus grand des mercis s'adresse à Thibault, pour son soutien quotidien depuis le premier jour. Merci de m'avoir toujours encouragée à persévérer et à donner le meilleur de moi-même. Ton courage et ta détermination sont pour moi une véritable source d'inspiration.

Contents

Introduction 1

Cancer is one of the major health challenges of our time, and it remains a central focus of medical research. To support progress in this field, the development of advanced analytical methods is essential. They have to better reflect the realities of clinical studies and help improve the evaluation of new treatments.

Clinical trials play a key role in assessing the effectiveness of innovative therapies. In broad terms, these trials include groups of patients affected by the same type of cancer or disease. In phase III trials, patients are randomly assigned to receive either the experimental treatment or the current standard of care. Each patient is then followed throughout the treatment period and beyond, during a dedicated follow-up. To assess the benefits of the therapy, a variety of measurements are collected. Typically, the occurrence of key clinical events, such as death or disease progression, and the time at which they occurred are recorded. These constitute the ***survival data***. In parallel, other patient-level measurements are repeatedly collected over time, for instance biomarkers or clinical assessments. These form the ***longitudinal data***. As we will see, jointly analysing these two types of information, rather than considering them separately, is crucial for understanding the relationship between longitudinal patient trajectories and survival events.

It is also well recognised that cancer and associated treatments affect more than survival alone. Together with the side effects of its treatment, cancer also impacts the physical, psychological, and even, social well-being of patients. For this reason, clinical trials increasingly seek to complement traditional clinical endpoints with the patient's own perspective of their health. These measurements, broadly referred to as patient-reported outcomes (PROs) often include ***health-related quality of life (HRQoL)*** and are often collected longitudinally, capturing the evolution of patients' perceptions throughout their treatment journey. Understanding how best to analyse HRQoL data, on their own and jointly with survival outcomes, has become an essential aspect of recent clinical research and appropriate analytic methods have to be employed (Fayers and Machin, 2013; Bottomley et al., 2016).

The remainder of this chapter provides a general introduction to these topics: the joint analysis of survival and longitudinal data, and the analysis of HRQoL data. It also defines the thesis objectives and outlines the structure of the following chapters.

1.1 Joint analysis of survival and longitudinal Data

1.1.1 Mixed-Effects Models for Longitudinal Data

Longitudinal data consist of repeated measurements collected on the same individuals at different time points. This structure provides information on how each subject evolves over time and allows within-subject and between-subject variability to be studied.

A standard modelling framework for analysing longitudinal responses is the class of mixed-effects models, also referred to as linear or generalized linear mixed models depending on the distribution of the outcome. These models include both fixed effects, which describe the average evolution at the population level, and random effects, which capture subject-specific deviations from this population trajectory (Verbeke and Molenberghs, 2000).

Let $y_{ij} = y_i(t_{ij})$ denote one longitudinal measurement for subject i ($i = 1, \ldots, n$) recorded at time t_{ij} ($j = 1, \ldots, n_i$). A generalized linear mixed model (GLMM) is defined through the relation

$$g\{\mu_i(t)\} = \mathbf{x}_i(t)^\top \boldsymbol{\beta} + \mathbf{z}_i(t)^\top \mathbf{b}_i, \qquad \mu_i(t) = \mathbb{E}\{y_i(t) \mid \mathbf{b}_i\},$$

where $g(\cdot)$ is a known link function, $\mathbf{x}_i(t)$ and $\mathbf{z}_i(t)$ are the design vectors for the fixed and random effects, and $\boldsymbol{\beta}$ denotes the vector of fixed-effects coefficients. The vector $\mathbf{b}_i$ contains the random effects for subject i, which are typically assumed to follow a multivariate normal distribution with mean zero and covariance matrix $\mathbf{D}$. Conditional on the random effects $\mathbf{b}_i$, the distribution of $y_i(t)$ is assumed to belong to the exponential family, including in particular Gaussian, binomial or Poisson distributions. For continuous Gaussian responses, we have the identity link $g(\cdot) = I(\cdot)$, and the observed measurements are often described as,

$$y_i(t) = \mu_i(t) + \varepsilon_i(t),$$

where the error term $\varepsilon_i(t)$ is assumed independent of $\mathbf{b}_i$ and normally distributed with variance σ^2_ε. The longitudinal dependence between repeated measurements is induced through the subject-specific random effects $\mathbf{b}_i$ in $\mu_i(t)$.

A major challenge in the analysis of longitudinal outcomes is that these measurements are frequently incomplete. Although patients are usually scheduled to attend visits at predetermined time points, in practice it is common that some visits are missed for a variety of reasons. Missing values in longitudinal studies typically arise in two different ways. The first situation corresponds to intermittent missingness, where some measurements are missing but later observations are still available. The second situation occurs when no further measurements are recorded after a certain time point, in which case the subject is considered to have dropped out of the study.

Intermittent missingness is often handled under the assumption that the missing data mechanism is missing at random (MAR) in the sense of Little and Rubin (2002). In this setting, mixed-effects models accommodate irregular and subject-specific measurement times, and estimation remains valid under the MAR assumption (Verbeke and Molenberghs, 2000).

In contrast, dropout may be informative, for example when the dropout process is related to the underlying health status of the patient, such as in cases of disease progression or death. This may correspond to a missing not at random (MNAR) mechanism. Ignoring this dependency may lead to biased estimates of the longitudinal process. To appropriately account for informative dropout, the longitudinal response and the dropout mechanism could be modelled jointly (Wulfsohn and Tsiatis, 1997).

1.1.2 Proportional Hazards Models for Survival Data

Survival analysis is concerned with modelling the time until a certain event of interest occurs. For each subject i, the true event time T_i^* may be unobserved due to censoring. In this work, we focus on right censoring, which occurs when the subject drops out before experiencing the event or when the study ends before the event is observed. This is indicated by the censoring indicator δ_i, which is 0 if the observation is censored and 1 if the event is observed. The observed time-to-event is $T_i = \min(T_i^*, C_i)$, where C_i denotes the censoring time of subject i.

The Cox proportional hazards (PH) model (Cox, 1972) is one of the most widely used models to analyse survival data. It is defined through the hazard function $h_i(t)$, representing the instantaneous event risk at time t for subjects still event-free. For subject i, we write

$$h_i(t) = \lim_{dt \to 0} \frac{\Pr(t \leq T_i^* < t + dt \mid T_i^* \geq t)}{dt} = h_0(t) \exp\left(\boldsymbol{\gamma}^\top \mathbf{w}_i\right), \qquad t > 0, \tag{1.1}$$

where $\mathbf{w}_i$ denotes a vector of baseline covariates, $\boldsymbol{\gamma}$ the corresponding coefficients, and $h_0(t)$ the baseline hazard. This formulation implicitly assumes that covariates do not vary over follow-up.

However, many studies aim to assess whether the event risk is influenced by covariates that change over time. Andersen and Gill (1982) proposed an extension of the Cox model that accommodates time-dependent covariates $y_i(t)$,

$$h_i(t) = h_0(t) \exp\{\boldsymbol{\gamma}^\top \mathbf{w}_i + \boldsymbol{\alpha}^\top y_i(t)\}.$$

Although this extended model provides a simple approach to incorporate time-varying information, its use requires great care. Time-dependent covariates can be classified into two types: *exogenous* and *endogenous* covariates. Broadly speaking, exogenous covariates are processes whose evolution is not affected by the occurrence of the event. A typical example is an external exposure such as air pollution. In contrast, covariates measured on the patient, such as biomarkers or clinical parameters, are endogenous: their values may be influenced by the health status of the patient and the occurrence of an event.

This distinction is crucial for extended Cox models, which assume that time-dependent covariates are predictable processes and measured without error. In the counting-process formulation, this means that the value of a covariate at time t must be known based only on the information available strictly before t. These requirements are unrealistic for endogenous covariates, whose evolution is driven by the underlying health status and may be influenced by the event process. The extended Cox model is therefore only appropriate for exogenous time-dependent covariates.

In clinical studies, longitudinal markers are almost always endogenous. Ignoring their specific features and using them directly in an extended Cox model generally leads to biased estimates of the covariate effect (Rizopoulos, 2012a).

1.1.3 Joint Modelling of Longitudinal and Survival Data

When longitudinal measurements (e.g. biomarkers) and survival outcomes are collected on the same individuals, the two processes are often closely related. The risk of experiencing a clinical event may depend on the underlying evolution of the biomarker, while the observation of the biomarker itself may be stopped when the event occurs. In such settings, using an extended Cox model with a time-dependent covariate or analysing the longitudinal data

alone would ignore this mutual dependency and may lead to biased estimates.

Joint models provide an appropriate framework to analyse longitudinal measurements and time-to-event data simultaneously. By modelling both processes together, they appropriately account for the endogenous nature of longitudinal biomarkers, their measurement error, and the non-random dropout caused by the occurrence of the event. Joint models are particularly useful when the objectives are to (i) estimate the effect of an endogenous marker on the risk of an event, (ii) study the evolution of a biomarker in the presence of non-random dropout, or (iii) assess the association between the two processes.

The most widely used formulation is the shared random effects joint model (Wulfsohn and Tsiatis, 1997; Tsiatis and Davidian, 2004), in which the association between the two processes is captured through subject-specific random effects appearing in both submodels. An alternative approach is the latent class joint model (Proust-Lima et al., 2009), which assumes that the population is divided into distinct latent subgroups, each characterised by its own longitudinal pattern and event risk. In this work, we focus exclusively on shared random effects joint models.

Model specification

In this section, to formulate the joint model in its simplest form, we consider one Gaussian longitudinal response and one terminal event. The shared random effects joint model consists of two submodels: a linear mixed-effects model for the repeated longitudinal measurements and a survival model for the individual risk of the event of interest. These two submodels are linked through a function of the random effects, which is incorporated into the linear predictor of the survival model.

Using the previously defined notation, the basic joint model is written as

$$\begin{cases} y_i(t) = \mu_i(t) + \varepsilon_i(t), & \varepsilon_i(t) \sim \mathcal{N}(0, \sigma_\varepsilon^2), \\ \mu_i(t) = \mathbf{x}_i(t)^\top \boldsymbol{\beta} + \mathbf{z}_i(t)^\top \mathbf{b}_i, & \mathbf{b}_i \sim \mathcal{N}(0, \mathbf{D}), \\ h_i(t) = h_0(t)\, \exp\left\{\boldsymbol{\gamma}^\top \mathbf{w}_i + \boldsymbol{\alpha}^\top\, m_i(t, \mathbf{b}_i)\right\}, \end{cases} \tag{1.2}$$

Here, $\boldsymbol{\alpha}$ quantifies the strength of the association between the longitudinal and survival processes, and $m_i(\cdot)$ denotes the association structure. Common choices include:

- *Current value association*: $m_i(t, \mathbf{b}_i) = \mu_i(t)$, where the event risk at time t depends on the true underlying marker level at the same time.

- *Random effects association*: $m_i(t, \mathbf{b}_i) = \mathbf{b}_i$, where only the subject-specific random effects are included in the survival model, capturing individual deviations from the population-average trajectory.
- *Time-dependent slope association*: $m_i(t, \mathbf{b}_i) = \big(\mu_i(t), \mu_i'(t)\big)$, where the event risk at time t depends both on the true current value of the marker and on the slope at t.

Additional association structures are described in Rizopoulos (2012a).

With the Cox PH model presented in Section 1.1.2, it is common to leave the baseline hazard function $h_0(t)$ unspecified in order to avoid misspecification of the survival time distribution, leading to a semi-parametric model. However, within the joint modelling framework, this strategy may result in biased inference (Hsieh et al., 2006). To avoid these issues, $h_0(t)$ is typically specified explicitly. A standard option is to assume a parametric form (e.g. Weibull, log-normal, or Gamma). Alternatively, and often preferably, a flexible parametric specification of the baseline hazard can be adopted, such as piecewise-constant functions or regression splines (Rizopoulos, 2012a). In this work, we consider a penalized spline–based approach, which allows control of the smoothness of the baseline hazard while avoiding overfitting (Rondeau et al., 2007; Rizopoulos, 2016). We express the logarithm of the baseline hazard function as

$$\log\{h_0(t)\} = \sum_{u=1}^{U} \gamma_{h_0,u} B_u(t),$$

where $B_u(t)$ denotes the uth basis function of a B-spline defined on equidistant knots over $(0, \max_i T_i)$ and $\gamma_{h_0,u}$ is the associated regression coefficient. Smoothness is ensured through a penalty on the spline coefficients.

Finally, the key assumption of the joint model is conditional independence given the random effects, which account for both the association between the longitudinal and event outcomes and the correlation between repeated longitudinal measurements.

Model estimation

Early approaches to the estimation of joint models relied on ***two-stage methods*** (Tsiatis et al., 1995). In the first stage, the longitudinal submodel is fitted independently, providing estimates of the fixed effects $\hat{\boldsymbol{\beta}}$ and subject-specific random effects estimates $\hat{\mathbf{b}}_i$. In the second stage, these quantities are treated as known and inserted into the survival submodel via $m_i\big(t \mid \hat{\mathbf{b}}_i, \hat{\boldsymbol{\beta}}\big)$. Although computationally attractive, this approach ignores the non-random dropout

mechanism in the first stage and leads to biased estimates (Tsiatis and Davidian, 2004).

Full ***joint estimation methods*** were later proposed, enabling simultaneous estimation of all parameters through frequentist or Bayesian inference. Both rely on the joint likelihood constructed from the joint distribution of longitudinal and time-to-event outcomes. Under the conditional independence assumption introduced above, the joint likelihood is

$$p(T, \delta, y|\theta) = \prod_{i=1}^{n} \left[\int p(T_i, \delta_i|\mathbf{b}_i, \theta_s) \prod_{j=1}^{n_i} p(y_{ij} \mid \mathbf{b}_i, \theta_y)\; p(\mathbf{b}_i|\theta_b)\, d\mathbf{b}_i \right], \quad (1.3)$$

where $\theta = (\theta_s, \theta_y, \theta_b)$ is the full parameter vector. Let $p(\cdot)$ denote a generic probability density (or mass) function. The contributions of the survival and longitudinal submodels are given by

$$p(T_i, \delta_i \mid \mathbf{b}_i, \theta_s) = h_i(T_i \mid \theta_s, \mathbf{b}_i)^{\delta_i} \exp\left\{ - \int_0^{T_i} h_i(s \mid \theta_s, \mathbf{b}_i)\, ds \right\}, \quad (1.4)$$

$$p\big(y_{ij} \mid \theta_y, \mathbf{b}_i\big) = (2\pi\sigma_\varepsilon^2)^{-1/2} \exp\left(- \frac{\big(y_{ij} - \mu_i(t_{ij})\big)^2}{2\sigma_\varepsilon^2} \right).$$

with $p(\mathbf{b}_i \mid \theta_b)$ assumed multivariate normal.

The integral over time in the definition of the survival function in Equation 1.4 generally does not have a closed-form solution and must therefore be evaluated numerically, for example using Gauss–Kronrod quadrature, when the association structure depends explicitly on time. In contrast, simpler random-effects–based association structures may lead to closed-form expressions and substantially simpler computations.

In the frequentist framework, joint models are typically estimated by maximizing the likelihood, most commonly using the EM algorithm (Wulfsohn and Tsiatis, 1997), Newton-type optimization (Thiébaut et al., 2005; Crowther et al., 2016), or maximum penalized likelihood approaches (Rondeau et al., 2007). EM algorithms provide stable estimation but may converge slowly, whereas Newton-like and penalized likelihood methods are often faster. However, these approaches generally rely on numerical integration over the random effects, which can become computationally challenging in models with complex or high-dimensional random-effect structures.

Under the Bayesian framework, parameters are estimated by sampling from the posterior distribution using Markov chain Monte Carlo or Hamiltonian

Monte Carlo. Consequently, numerically solving the integral over the random effects in Equation 1.3 is no longer required, unlike in the frequentist approach. The posterior distribution satisfies

$$p(\theta, \mathbf{b}) \propto \prod_{i=1}^{n} p(T_i, \delta_i \mid \mathbf{b}_i, \theta_s) \prod_{j=1}^{n_i} p(y_{ij} \mid \mathbf{b}_i, \theta_y)\ p(\mathbf{b}_i \mid \theta_b)\ p(\theta),$$

where $p(\theta)$ denotes the prior distribution.

For simple joint models, Bayesian computation may be slower than frequentist methods. However, for more challenging models (e.g. non-Gaussian longitudinal outcomes, multiple longitudinal markers, or complex association structures), the speed of the Bayesian approach becomes comparable or even faster (Rizopoulos, 2016). Moreover, Bayesian approaches become particularly attractive in these complex models, as all model components are estimated jointly through posterior sampling. Uncertainty is naturally propagated across the different parts of the model, and explicit high-dimensional numerical integration over the random effects can often be avoided.

Available packages

Several R packages have been developed for fitting joint models, each offering different levels of flexibility and computational strategies. The package `JM` (Rizopoulos, 2010) was the first widely used implementation based on maximum likelihood and the EM algorithm, while its Bayesian extension `JMbayes` (Rizopoulos, 2016), and the more recent `JMbayes2`, provides greater modelling flexibility, including multivariate longitudinal outcomes, competing risks, and a wide range of association structures. The `joineR` and `joineRML` packages implement frequentist EM and Monte Carlo EM algorithms, respectively, with `joineRML` supporting multivariate longitudinal processes. The package `frailtypack` (Rondeau et al., 2012) provides semiparametric penalized likelihood estimation for joint models. In particular, it includes joint models for longitudinal data and a terminal event (`longiPenal`) and trivariate joint models for longitudinal data, recurrent events, and a terminal event (`trivPenal`). Bayesian approaches are also available through `rstanarm`, which uses Hamiltonian Monte Carlo, and `bamlss`, which fits flexible additive joint models. More recently, `INLAjoint` has introduced fast Bayesian joint modelling using the Integrated Nested Laplace Approximation (INLA). Finally, the package `lcmm` (Proust-Lima et al., 2017) provides tools for estimating latent class joint models under the frequentist framework.

1.2 Analysis of Health-Related Quality of Life Data

1.2.1 Context : Global Cancer Burden

Cancer is one of the leading public health challenges: in 2022, nearly 20 million new cancer cases were diagnosed worldwide. Demographic trends suggest that this number could climb to 35 million by 2050, largely driven by population ageing, population growth, and changes in exposure to major risk factors (WHO, 2024).

At the same time, cancer research has progressed substantially over recent decades, leading to advances in early detection, treatment modalities, supportive care, and follow-up. These improvements have significantly increased the number of cancer survivors. For instance, according to NIH (2025), the five-year relative survival rate for all cancers combined rose from 65.8% for individuals diagnosed in 2000 to 72.5% for those diagnosed in 2017. This growing population of survivors underscores the importance of not only prolonging life, but also ensuring that the additional years are lived with an acceptable quality of life. In contrast, for cancers with poor prognosis and limited survival, improving patients' quality of life during the course of the disease becomes a key objective.

Indeed, beyond survival, cancer affects patients physically, psychologically, and socially. Cancer treatments are known to be invasive and associated with numerous side effects (e.g. alopecia, vomiting, fatigue). It is therefore desirable that medical progress focuses not only on extending survival, but also on improving the health-related quality of life of patients.

1.2.2 HRQoL criteria in clinical trials

Clinical trials play a key role in the development of innovative therapies and follow a phased structure: Phase I trials assess safety, tolerability, and dose ranges in small groups; Phase II trials investigate preliminary efficacy and optimal dosing; and Phase III trials involve large patient populations to confirm therapeutic benefit, monitor adverse events, and compare the experimental treatment with existing standard of care.

A fundamental aspect of any clinical trial is the choice of endpoints, which are events or outcomes measured to determine whether the therapy under investigation is beneficial. In oncology, the most widely accepted primary endpoints include:

- **Overall survival (OS):** time from randomisation to death.

- **Progression-free survival (PFS):** time from randomisation until the first evidence of disease progression or death.

These endpoints capture essential aspects of disease control and patient prognosis. However, they may not fully represent the global impact of cancer and its treatments on patients' functioning and well-being. This is particularly relevant in oncology, where treatment side effects and symptom burden can significantly affect patients' daily lives, even when clinical outcomes appear favorable.

This has motivated the increasing use of complementary endpoints capable of capturing the patient's perspective. Health-related Quality of Life (HRQoL), a patient-reported outcome (PRO), has emerged as a key component in the evaluation of therapeutic benefit.

1.2.3 Assessing HRQoL

Health-Related Quality of Life (HRQoL) is a multidimensional construct that captures the patient's own perception of how illness and its treatment affect symptoms, functional abilities, psychological well-being, and overall health (Fayers and Machin, 2013). Because HRQoL evolves over time, it is typically collected longitudinally during treatment and follow-up. As such, HRQoL constitutes a rich and challenging endpoint, requiring careful conceptualization and appropriate methodological tools for its evaluation.

HRQoL questionnaires

Because HRQoL is not directly observable, it is assessed using validated self-reported questionnaires. These instruments are typically designed to measure multiple dimensions of quality of life. Their use provides a standardized, reproducible, and comparable way to report how patients actually experience their illness and its treatment.

Questionnaires developed by the European Organisation for Research and Treatment of Cancer (EORTC) are among the most widely used. In particular, the core questionnaire EORTC QLQ-C30 (Aaronson et al., 1993) is the reference tool for assessing HRQoL in international cancer clinical trials. Originally published in 1993, it has been extensively developed and rigorously validated to accurately represent HRQoL in patients with various cancer types and across different cultures.

The QLQ-C30 contains 30 questions (items) which assess distinct HRQoL dimensions (scales). There are both multi-item scales and single-item measures.

These include:

- five functional scales (physical, role, emotional, cognitive, and social functioning),
- three symptom scales (fatigue, pain, nausea/vomiting),
- six single items (dyspnoea, insomnia, appetite loss, constipation, diarrhoea, financial difficulties),
- and a global health status / overall quality of life scale.

For the first 28 items, responses are given on a four-point Likert scale ranging from "Not at all" to "Very much". The two items reflecting global health and overall quality of life use a seven-point scale, allowing patients to express a broader range of perceptions.

In addition to the core questionnaire, the EORTC group has developed numerous cancer-site-specific modules that can be administered alongside the QLQ-C30. For example, the EORTC QLQ-BN20 is a 20-item module designed specifically for patients with brain tumours. It assesses domains such as visual, motor or communication difficulties, which are particularly relevant for this tumour type.

Classical Scoring Approach

HRQoL questionnaires are typically analysed through a scoring procedure that aggregates item responses into interpretable scale scores. These scores are constructed as the mean of the item responses and then standardised so that all dimensions are expressed on a common 0–100 scale. A higher score represents a healthier level of HRQoL (Fayers et al., 2001).

Let a scale be composed of K items, each with responses taking values in $\{1, \ldots, L\}$, where higher categories indicate a greater intensity of the dimension being measured. We denote the responses for a given individual i by $(y_{i1}, \ldots, y_{iK})$. For symptom scales, where higher scores correspond to more severe symptoms, the standardised score S_i is obtained through the linear transformation:

$$S_i = \left(\frac{\frac{1}{K} \sum_{k=1}^{K} y_{ik} - 1}{L - 1} \right) \times 100. \tag{1.5}$$

For functional scales, the transformation is reversed so that higher scores represent a healthier level of functioning:

$$S_i = \left[1 - \frac{\frac{1}{K} \sum_{k=1}^{K} y_{ik} - 1}{L - 1} \right] \times 100. \tag{1.6}$$

In the presence of missing data, EORTC guidelines recommend scoring a scale if at least half of its items have been answered.

Limitations of HRQoL Scores

Although this approach provides a simple method for summarizing patient-reported outcomes, it presents several methodological limitations, especially when used in longitudinal analyses (Tennant and Conaghan, 2007; Gorter et al., 2015).

In most applications, the scale scores are treated as continuous variables and analysed using linear mixed models (LMM). This approach implicitly assumes that the scores, obtained by averaging ordinal item responses handled as if they were integers, approximately follow a Gaussian distribution. However, the bounded nature of the score introduces potential ceiling and floor effects, leading to asymmetry and non-negligible deviation from normality. Moreover, these scores do not account for the ordinal nature of the underlying data. They assume equal distances between adjacent response categories (e.g. between "a little" and "quite a bit"), an assumption that may not hold in practice. In addition, individuals with different item-response patterns may obtain the same total score, resulting in an important loss of information.

Therefore, classical LMMs applied to HRQoL scores may fail to capture the true complexity of the underlying latent construct and may yield biased conclusions.

Motivated by these limitations, alternative approaches have been explored, including models that work directly with item-level (raw) responses instead of aggregated scores, such as IRT-based methods.

Item response theory approach

Item Response Theory (IRT) provides a psychometric framework for modelling the relationship between a patient's latent health status and the probability of endorsing specific questionnaire responses, while explicitly accounting for the ordinal nature of the data (Embretson and Reise, 2013). Unlike classical scoring methods, which treat all items as equally informative and assume linear response scales, IRT characterizes item behaviour through parameters that reflect their difficulty and discriminative ability.

IRT assumes that each respondent possesses an unobserved (latent) level of the construct of interest (e.g., physical functioning or fatigue), denoted by η.

This latent trait is derived from the pattern of item responses and is represented on a continuous scale whose extremes correspond to the lowest and highest levels of the construct.

Given this framework, the probability of selecting a particular response category for item k is modelled as a function of η and a set of item-specific parameters. For ordinal response scales, as used in most HRQoL instruments, the most commonly applied polytomous IRT model is the graded response model (GRM) (Samejima, 1969).

Formally, the GRM defines the cumulative probability of selecting category l or below for a given item. More precisely, for patient i and item k,

$$\Pr(y_{ik}(t) \leq l \mid \eta_i(t)) = \frac{1}{1 + \exp\{-(d_{kl} - a_k\, \eta_i(t))\}}.$$

where a_k is the discrimination parameter and d_{kl} are the ordered category thresholds. The discrimination parameter quantifies to what extent the item distinguishes between patients with different values of $\eta_i(t)$, while the thresholds identify the points on the latent continuum at which the probability of endorsing category l or below equals 0.5. Items with a large a_k value provide greater measurement precision, and thresholds d_{kl} that are sufficiently spread along the continuum ensure that the item remains informative across a broad range of patient severities. The probability of selecting an exact category is obtained by subtracting adjacent cumulative probabilities.

Overall, IRT provides a more rigorous and interpretable framework for analysing HRQoL questionnaires, offering deeper insight into how different items reflect patient-reported health levels and improving measurement precision compared with classical scoring methods (Hays et al., 2000).

Underlying assumptions of IRT models

Like any statistical model, IRT models rely on several key assumptions (Embretson and Reise, 2013).

First, *unidimensionality* assumes that the set of items within a given scale measures a single underlying construct. This assumption is essential, as it defines the meaning of the latent variable and ensures that it provides a coherent summary of the item responses.

Second, the assumption of *local independence* states that, conditional on the latent trait, item responses are independent. In other words, the latent variable is assumed to capture all the dependence between items.

Third, *monotonicity* assumes that higher levels of the latent trait are associated with a higher probability of endorsing response categories reflecting a greater level of the underlying construct. This property ensures that the ordering of response categories is meaningful and consistent with the interpretation of the latent variable.

In addition, IRT models rely on a *measurement invariance* assumption, according to which the relationship between the latent trait and the item responses remains stable across individuals and over time (Mellenbergh, 1989). In the context of HRQoL data, this assumption may be challenged by differential item functioning (DIF), where items behave differently across patient subgroups (Holland and Wainer, 2012), or by response shift, where patients' perception or interpretation of questionnaire items changes over time (Sprangers and Schwartz, 1999).

1.3 Objectives and outline of the thesis

The objective of this thesis is to develop and extend joint modelling frameworks for the analysis of longitudinal outcomes and time-to-event data, motivated by challenges arising in current clinical trials. Particular attention is devoted to complex longitudinal structures, especially patient-reported HRQoL data. The proposed methodologies aim to be both statistically reliable and practically applicable for joint modelling in oncology research.

The remainder of this thesis is organised into three methodological chapters, each addressing a specific challenge in the joint analysis of longitudinal and time-to-event data. All chapters are illustrated using data from a clinical trial conducted in patients with recurrent glioblastoma.

Chapter 2 develops an extension of classical joint models, FlexibleJM, which relaxes the assumption of linear covariate effects in the survival submodel. Using Bayesian penalised B-splines, the proposed framework accommodates nonlinear associations while allowing for non-Gaussian longitudinal responses through a generalized linear mixed model formulation.

Chapter 3 introduces JMIRT, a joint modelling strategy specifically designed for multivariate ordinal HRQoL data in the presence of disease-related dropout. Instead of treating longitudinal measurements as predictors in the survival component, the approach focuses on the latent HRQoL process and incorporates dropout-related risks directly into its evolution.

Chapter 4 investigates the use of two-stage approaches as a computation-

ally efficient alternative to full joint estimation. More precisely, we propose a SLOPE-CORRECTED TWO-STAGE approach that mitigates the biases typically observed in standard two-stage procedures. This methodology is applied to multidimensional latent trait joint models, enabling the joint analysis of multiple HRQoL domains while retaining substantial computational advantages.

Chapter 5 concludes the thesis by summarising the main methodological contributions, discussing their limitations, and outlining perspectives for future research in joint modelling of longitudinal and time-to-event data, with a particular focus on HRQoL outcomes in oncology.

Flexible survival submodel in joint models

2

In medical studies, repeated measurements of biomarkers and time-to-event data are often collected during follow-up. To assess their association, joint models are widely used, typically combining a linear mixed model for the longitudinal process with a proportional hazards model for survival, which assumes linear covariate effects on the log-hazard. In this chapter, we extend this framework by allowing nonlinear effects of baseline survival covariates through Bayesian P-splines (penalised B-splines), while also modelling the baseline hazard flexibly. The approach accommodates non-Gaussian longitudinal outcomes via a generalized linear mixed model. We illustrate the method using data from patients with first progression of glioblastoma.

The material presented in this chapter is based on the paper: Doms H, Lambert P, Legrand C. (2024). Flexible joint model for time-to-event and non-Gaussian longitudinal outcomes. *Statistical Methods in Medical Research*. 33(10), 1783–1799

2.1 Introduction

In medical studies for which the primary interest is to record the time at which a particular event occurs, information on multiple biomarkers is also often collected longitudinally throughout the follow-up period. This provides a combination of survival and longitudinal information about each individual under study. Many methods allow the longitudinal and time-to-event responses to be studied separately, but it is often preferable to analyse them jointly (Rizopoulos, 2012a). Indeed, longitudinal measurements could have a predictive role in the analysis of patient survival. A more suitable objective is therefore to measure the association between longitudinal measurements and the event risk. In addition, it is important to take into account that the longitudinal response is typically measured with error and has missing data.

To meet this need, a class of statistical models has been developed: joint models for longitudinal and time-to-event data (Faucett and Thomas, 1996). The basic idea behind the joint model approach consists of three steps : first, define a model for the longitudinal biomarker measurements; second, define a model for the time-to-event data; and third, link the two submodels with a shared latent association structure. We consider the survival outcome as the primary interest and we incorporate a summary measure of the longitudinal process as a covariate in the time-to-event model.

The seminal formulation of this model was introduced by Faucett and Thomas (1996) and Wulfsohn and Tsiatis (1997) and is based on a single longitudinal outcome assumed to be normally distributed and a single time-to-event outcome that are linked by adding the expected values of the longitudinal response as a covariate in the event risk model. The joint model framework has subsequently been widely studied in the literature and several extensions have been proposed. These include, among others, non-Gaussian longitudinal outcome (Viviani et al., 2014; Rizopoulos, 2016), multiple longitudinal outcomes (Brown et al., 2005; Chi and Ibrahim, 2006; Rizopoulos and Ghosh, 2011) and multiple recurrent or competing events (Elashoff et al., 2008; Andrinopoulou et al., 2014). Chi and Ibrahim (2006) presented a joint model considering multiple longitudinal and multiple survival outcomes. Complete overviews of joint modelling techniques have been published (Rizopoulos, 2012a; Papageorgiou et al., 2019). Another important extension is to allow greater flexibility in joint modelling. Indeed, real data sets often exhibit nonlinear trajectories or effects. Rizopoulos and Ghosh (2011) propose a spline-based approach to describe subject-specific nonlinear longitudinal evolutions. Andrinopoulou et al. (2017) describe a Bayesian joint model that allows a time-varying coefficient to link the longitudinal and survival submodels, using P-splines. Similarly, Köhler et al. (2017) propose a flexible additive joint model for a continuous longitudinal outcome and single time-to-event outcome under the Bayesian framework using P-splines. This model is highly flexible as it is able to represent all the components of the joint model in an additive and flexible way. However, this approach can only accommodate a Gaussian longitudinal marker.

In this chapter, we focus on flexible specifications of the survival part in the joint model and more specifically on the potential nonlinearity of the effect of the baseline survival covariates. To our knowledge, no work has focused on assessing whether accounting for nonlinear effects of survival covariates could improve the fit of a joint model while considering a non-Gaussian longitudinal response. We assume nonlinear effects by using Bayesian P-splines (Lang and Brezger, 2004). We also use P-splines to describe the logarithm of

the baseline hazard. To force smoothness of the fitted curves and to avoid overfitting, a roughness penalty is applied. Our method is valid for Gaussian and non-Gaussian distributions of the longitudinal marker.

The study that motivated our research concerns patients with a first progression of glioblastoma. Glioblastoma is the most common brain cancer in adults, but its survival prognosis is very poor. Therefore, it is essential to be able to identify prognostic factors for this cancer in order to guide treatment strategies. We aim to use our methodology on these data to assess the effects of potential prognostic factors more accurately.

The rest of this chapter is organised as follows. In Section 2.2, we describe the formulation of the joint model and introduce our proposed flexible survival submodel using P-splines. In Section 2.3, we present the Bayesian estimation of the model. In Section 2.4, we discuss a simulation study to evaluate the statistical performance of our approach. Section 2.5 illustrates our approach to glioblastoma data. Finally, in Section 2.6, we conclude the chapter with a discussion.

2.2 The joint model

2.2.1 Longitudinal submodel

For every individual ($i = 1, \ldots, n$) we collect longitudinal response values $y_{ij} = y_i(t_{ij})$ observed at time t_{ij} ($j = 1, \ldots, n_i$). The subject-specific evolution over time of the longitudinal response is modeled using a generalized mixed effect model (GLMM) (Brown and Prescott, 2006). In particular, we postulate

$$\begin{cases} g\{\mathbb{E}\{y_i(t) \mid \mathbf{b}_i\}\} = \mathbf{x}_i(t)^\top \boldsymbol{\beta} + \mathbf{z}_i(t)^\top \mathbf{b}_i \\ \mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}), \end{cases} \tag{2.1}$$

where $\boldsymbol{\beta}$ denotes the p-dimensional vector of fixed effects associated with the time-dependent design vector $\mathbf{x}_i(t)$ and $\mathbf{b}_i$ denotes the q-dimensional vector of random effects associated with $\mathbf{z}_i(t)$. The random effects $\mathbf{b}_i$ allow each of the subjects to have a different response profile in time, and we assume a multivariate normal distribution with zero mean and variance-covariance matrix $\mathbf{D}$. Conditional on the random effects $\mathbf{b}_i$, the distribution of $\mathbf{y}_i$ is assumed to be a member of the exponential family, with $g(.)$ denoting a known link function. When the observed measurements on individuals derive from a counting process, a common choice is to use a logarithm link $g(.) = \log(.)$, which is the canonical link function for the Poisson distribution. When the observed measurements have a binomial distribution, the logit can be used as the canonical link function. In the simple case of normally distributed longitudinal data, the associated canonical link function is the identity function.

2.2.2 Survival submodel

Let T_i^* denote the true event time for the i-th individual, C_i the right-censoring time, $\delta_i = I(T_i^* \leq C_i)$ the event indicator (1 if the ith individual experiences the event, 0 otherwise) and $T_i = \min(T_i^*, C_i)$ the reported time value. To link the longitudinal and survival outcomes, a common approach in the joint modelling framework assumes a proportional hazards model for the conditional hazard:

$$h_i(t \mid \mathcal{M}_i(t), \mathbf{w}_i) = h_0(t) \exp\Big(\boldsymbol{\gamma}_w^\top \mathbf{w}_i + m\{\mathcal{M}_i(t), \mathbf{b}_i, \alpha\}\Big), \qquad t > 0 \qquad (2.2)$$

where $h_0(t)$ denotes the baseline hazard function and $\mathbf{w}_i = [w_{i1}, \ldots, w_{iL}]^\top$ is the vector of baseline survival covariates for subject i with regression coefficients $\boldsymbol{\gamma}_w$. The covariate vectors $\mathbf{x}_i$ and $\mathbf{w}_i$ may overlap, so that some covariates can simultaneously influence the longitudinal and survival submodels (Crowther et al., 2013b). Within the joint modelling framework, careful specification of the baseline hazard function is required. In particular, $h_0(t)$ should not be left completely unspecified (Rizopoulos, 2012a). In this work, we specify $h_0(t)$ flexibly using spline-based representations. Finally, let $\mathcal{M}_i(t) = \{\eta_i(s), 0 \leq s < t\}$ denote the history of the underlying unobserved longitudinal process up to time t. Function $m\{\mathcal{M}_i(t), \mathbf{b}_i, \alpha\}$ specifies the association structure between the two submodels with parameter α tuning the strength of the association between the longitudinal process and the survival outcome. Various specifications are used in the literature (Rizopoulos, 2012a; Papageorgiou et al., 2019). In this article, we consider the current-value association structure defined by

$$m\{\mathcal{M}_i(t), \mathbf{b}_i, \alpha\} = \alpha\, \eta_i(t).$$

This parameterization is very intuitive and easy to interpret with $\exp(\alpha)$ giving the relative risk increase of an event at time t resulting from a simultaneous one unit increase in $\eta_i(t)$.

2.2.3 Flexible survival submodel

Model (2.2) assumes that all survival covariates have a linear effect on the log-hazard. We propose an extended version of this model that allows for possible nonlinear effects of some survival covariates. Specifically, we have

$$h_i(t \mid \mathcal{M}_i(t), \mathbf{w}_i, \mathbf{v}_i) = h_0(t) \exp\{\boldsymbol{\gamma}_w^\top \mathbf{w}_i + \sum_{j=1}^{J} f_j(v_{ij}) + \alpha \eta_i(t)\}, \qquad t > 0 \quad (2.3)$$

where covariates $\mathbf{v}_i = [v_{i1}, \ldots, v_{iJ}]^\top$ enter the log-hazard in a nonlinear way and $\mathbf{w}_i = [w_{i1}, \ldots, w_{iL}]^\top$ is a vector of categorical or additional continuous covariates with a corresponding vector of regression coefficients $\boldsymbol{\gamma}_w^\top =$

$[\gamma_{w1}, \ldots, \gamma_{wL}]$. The additive terms $\mathbf{f} = [f_1, \ldots, f_J]^\top$ are assumed to be smooth and represent nonlinear effects of the associated continuous covariates. In practice, exploratory analyses can be conducted to investigate potential nonlinear effects (see Appendix A1).

The corresponding survival function is defined by

$$S_i(t \mid \mathcal{N}_i(t), \mathbf{w}_i, \mathbf{v}_i) = \exp\Big(- \int_0^t h_0(s) \exp\{\boldsymbol{\gamma}_w^\top \mathbf{w}_i + \sum_{j=1}^{J} f_j(v_{ij}) + \alpha \eta_i(s)\}\, ds\Big).$$

A well-known approach used for modelling and estimating smooth functions relies on B-splines. A B-spline of degree d consists of $d+1$ polynomial pieces, each of degree d, that join together at d specific points called inner knots (Eilers and Marx, 1996). The additive terms f_j are expressed as

$$f_j(v_{ij}) = \sum_{k=1}^{K} \gamma_{v,jk} B_{jk}(v_{ij}),$$

where $\gamma_{v,jk}$ represents the B-spline coefficient associated to the B-spline basis function $B_{jk}(.)$ at the respective covariate value v_{ij}. For every f_j, we assume the same number K of basis functions $B_{jk}(.)$ for simplicity (associated to equidistant knots on the range of the jth covariate) and we define the $n \times K$ matrix $\mathbf{B}_j$ for which the element at the ith row and kth column correspond to $B_{jk}(v_{ij})$.

As introduced in Section 2.2.2 and suggested by Lambert and Eilers (2005) and by Rizopoulos (2012a) in the specific context of joint models, we use the same approach to model $h_0(t)$ and express the logarithm of the baseline hazard function as

$$\log\{h_0(t)\} = \sum_{u=1}^{U} \gamma_{h_0,u} B_u(t),$$

where $\gamma_{h_0,u}$ is the regression coefficient associated to the uth function of a B-spline basis associated to equidistant knots on $(0, \max_i T_i)$.

Choosing the optimal number and positions of knots is a complex task that has an impact on the estimated functions. Indeed, as the number of knots increases, the approximation becomes more flexible. However, too many knots can cause serious over-fitting. To overcome this problem, Eilers and Marx (1996) proposed a penalised B-splines approach, called P-splines. They suggest using a relatively large number of equally spaced knots and to counterbalance the flexibility of the fit by adding a roughness penalty based on differences of adjacent B-spline coefficients. As a result, the precise number

and location of knots become less critical, and a relatively large number of equally spaced knots can be safely used in practice.

P-splines can also be estimated in a Bayesian framework as Bayesian P-splines (Lang and Brezger, 2004; Jullion and Lambert, 2007). The difference penalties are replaced by their stochastic analogues i.e., random walk priors. We obtain the roughness penalty by imposing a Gaussian random field prior on the spline coefficients (Rue and Held, 2005) (see Section 2.3.2 for details). Both the log baseline hazard function and the additive terms $\mathbf{f}$ will be approximated using Bayesian P-splines.

Inference on the smooth components can be performed using the approximate hypothesis tests proposed by Wood (2013). These tests aim to assess whether a smooth term contributes to the predictor by testing the null hypothesis $H_0 : f_j = 0$. The approach relies on the fact that the estimated smooth, evaluated at the observed covariate values and denoted by $\hat{\mathbf{f}}_j$, can be approximated as Gaussian with mean $\mathbf{f}_j$ and covariance matrix $\mathbf{V}_{f_j}$, which is derived from the Bayesian interpretation of the penalized spline model. This approximation leads to a Wald-type statistic constructed from $\hat{\mathbf{f}}_j$ and its covariance matrix. Under the null hypothesis, this statistic is approximately chi-square distributed, with degrees of freedom based on the effective degrees of freedom (EDF) of the smooth. The use of effective degrees of freedom reflects the amount of penalization and prevents overestimating the significance of highly constrained components.

It should be noted that this test evaluates the presence of an effect rather than its form. In particular, evidence of nonlinearity must be assessed separately. In practice, nonlinearity is commonly assessed using the effective degrees of freedom of the smooth (Wood, 2017; Hastie and Tibshirani, 1990). The EDF measures the complexity of a smooth term and can be interpreted as the number of parameters effectively used to represent the function. Values close to one indicate an approximately linear effect, while larger values reflect increasing flexibility. Finally, graphical inspection of the estimated smooth functions, together with their confidence intervals, is used to validate and interpret the nature of the effects. Although these approaches are heuristic rather than formal tests, they are widely used in practice.

2.3 Bayesian estimation

We apply a Bayesian approach where inference is based on the posterior of the model parameters. In particular, we estimate the parameters of the joint model (2.1) and (2.3) using Markov Chain Monte Carlo (MCMC) methods.

2.3.1 Likelihood function

In the framework of joint models for longitudinal and survival data, the likelihood is derived under the common assumption that the longitudinal and survival processes are independent conditionally on the random effects $\mathbf{b_i}$. Moreover, the longitudinal measurements $\mathbf{y}_i$ of each individual are assumed independent given the random effects. Formally we have (Rizopoulos, 2016),

$$\begin{aligned} p(T_i, \delta_i, \mathbf{y}_i \mid \mathbf{b}_i; \theta) &= p(T_i, \delta_i \mid \mathbf{b}_i; \theta)\; p(\mathbf{y}_i \mid \mathbf{b}_i; \theta) \\ p(\mathbf{y}_i \mid \mathbf{b}_i; \theta) &= \prod_{j=1}^{n_i} p(y_i(t_{ij}) \mid \mathbf{b}_i; \theta) \end{aligned} \tag{2.4}$$

where $\theta = (\theta_s, \theta_y, \theta_b)$ contains θ_s the parameters for the survival outcome, θ_y the parameters for the longitudinal outcome and θ_b the parameters of the random-effect covariance matrix. Function $p(.)$ generically denotes the corresponding probability or density function.

The expression of the marginal likelihood contribution for the i^{th} subject is written as

$$\begin{aligned} L_i(\theta; T_i, \delta_i, \mathbf{y}_i) &= \int p(T_i, \delta_i, \mathbf{y}_i, \mathbf{b}_i; \theta)\, d\mathbf{b}_i \\ &= \int p(T_i, \delta_i \mid \mathbf{b}_i; \theta_s, \boldsymbol{\beta}) \Big[\prod_{j=1}^{n_i} p\big(y_i(t_{ij}) \mid \mathbf{b}_i; \theta_y\big) \Big] p(\mathbf{b}_i; \theta_b)\, d\mathbf{b}_i \end{aligned}$$

The marginal log-likelihood function formulated for all subjects is obtained by $l(\theta) = \sum_{i=1}^{n} \log L_i(\theta; T_i, \delta_i, \mathbf{y}_i)$.

The extended expression for the marginal likelihood contribution of the ith

subject in the survival submodel is given by

$$
\begin{aligned}
p(T_i, \delta_i \mid \mathbf{b}_i; \theta_s, \boldsymbol{\beta}) &= \left[h_i(T_i \mid \mathcal{N}_i(T_i); \theta_s, \boldsymbol{\beta})\right]^{\delta_i} S_i(T_i \mid \mathcal{N}_i(T_i); \theta_s, \boldsymbol{\beta}) \\
&= \left[h_0(T_i) \exp\{\boldsymbol{\gamma}_w^\top \mathbf{w}_i + \sum_{j=1}^{J} f_j(v_{ij}) + \alpha\eta_i(T_i)\}\right]^{\delta_i} \\
&\quad \times \exp\left(-\int_0^{T_i} h_0(s) \exp\{\boldsymbol{\gamma}_w^\top \mathbf{w}_i + \sum_{j=1}^{J} f_j(v_{ij}) + \alpha\eta_i(s)\}\, ds\right) \\
&= \left[\exp\left\{\sum_{u=1}^{U} \gamma_{h_0,u} B_u(T_i) + \boldsymbol{\gamma}_w^\top \mathbf{w}_i + \sum_{j=1}^{J}\sum_{k=1}^{K} \gamma_{v,jk} B_{jk}(v_{ij}) + \alpha\eta_i(T_i)\right\}\right]^{\delta_i} \\
&\quad \times \exp\left[-\exp\left\{\boldsymbol{\gamma}_w^\top \mathbf{w}_i + \sum_{j=1}^{J}\sum_{k=1}^{K} \gamma_{v,jk} B_{jk}(v_{ij})\right\}\right. \\
&\quad \left. \times \int_0^{T_i} \exp\left\{\sum_{u=1}^{U} \gamma_{h_0,u} B_u(s) + \alpha\eta_i(s)\right\} ds\right]
\end{aligned}
\tag{2.5}
$$

The distribution of the outcomes $\mathbf{y}_i$ is assumed to belong to the exponential family of distributions and so the likelihood contribution of the longitudinal submodel takes the form

$$
p(y_{ij} \mid \mathbf{b}_i; \theta_y) = \exp\left\{\frac{y_{ij}\, \psi_{ij}(\mathbf{b}_i) - c\{\psi_{ij}(\mathbf{b}_i)\}}{a(\phi)} - d(y_{ij}, \phi)\right\} \tag{2.6}
$$

where $\psi_{ij}(\mathbf{b}_i)$ and ϕ are the natural and scale parameters respectively and $c(.)$ and $d(.)$ are known functions that identify the member of the exponential family. For one-parameter distributions, functions $a(\phi)$ and $d(y_{ij}, \phi)$ can be simplified to a and $d(y_{ij})$ respectively (Brown and Prescott, 2006). Table 2.1 presents different choices of these functions to cover the most common distributions.

Table 2.1: Expression for $a(.)$, $c(.)$ and $d(.)$ for members of the exponential family

Distribution	a	$c(\psi)$	$d(y)$
Bernoulli	1	$\log(1+\exp(\psi))$	0
Binomial	$1/n$	$\log(1+\exp(\psi))$	$\log[n!/((ny)!\,(n-ny)!)]$
Poisson	1	$\exp(\psi)$	$-\log(y!)$
	$a(\phi)$	$c(\psi)$	$d(y, a(\phi))$
Normal	ϕ	$\psi^2/2$	$-\frac{1}{2}\{y^2/\phi + \log(2\pi\phi)\}$

The normality assumption for the distribution of random effects implies that

$$p(\mathbf{b}_i; \theta_b) = (2\pi)^{-q/2} \det(\mathbf{D})^{-1/2} \exp\big(-\mathbf{b}_i^\top \mathbf{D}^{-1} \mathbf{b}_i / 2 \big) \tag{2.7}$$

where q denotes the dimension of the random effect vector.

A numerical integration method is required for the evaluation of the integral involved in the survival contribution in (2.5), since this integral does not admit a closed-form solution under the proposed model specification. The integral is approximated using a 15-point Gauss–Kronrod quadrature rule.

$$\int_0^{T_i} \exp\{\log h_0(s) + \alpha \eta_i(s)\}\, ds \approx \frac{T_i}{2} \sum_{k=1}^{15} w_k \exp\{\log h_0(t_{ik}) + \alpha \eta_i(t_{ik})\}$$

where t_{ik} are the Gauss–Kronrod quadrature nodes on $[-1, 1]$ re-scaled to the interval $[0, T_i]$, and w_k are the corresponding Gauss–Kronrod quadrature weights.

2.3.2 Priors and posterior distributions

Bayesian estimation of model parameters is based on the posterior distribution. The expression for the posterior distribution of the model parameters is derived under the assumptions defined in (2.4). Using Bayes theorem, we obtain

$$p(\theta, \mathbf{b} \mid \mathbf{T}, \delta, \mathbf{y}) \;\propto\; \prod_{i=1}^{n} p(T_i, \delta_i | \mathbf{b}_i; \theta_s) \prod_{j=1}^{n_i} p\big(y_i(t_{ij}) | \mathbf{b}_i; \theta_y\big)\, p(\mathbf{b}_i; \theta_b)\; p(\theta) \tag{2.8}$$

where θ denotes the full parameter set with joint prior $p(\theta)$. The expressions for $p(T_i, \delta_i \mid \mathbf{b}_i; \theta)$, $p(y_i(t_{ij}) \mid \mathbf{b}_i; \theta)$ and $p(\mathbf{b}_i; \theta_b)$ can be found in (2.5), (2.6) and (2.7), respectively.

We use standard prior distributions for θ (Papageorgiou et al., 2019; Alsefri et al., 2020). In particular, independent univariate weakly informative normal priors are assigned to $\boldsymbol{\beta}$ (the vector of fixed effects in the longitudinal submodel), to $\boldsymbol{\gamma}_w$ (the vector of regression coefficients in the survival submodel), and to α (the association parameter). For the random-effects covariance matrix $\mathbf{D}$, we adopt a separation strategy in which $\mathbf{D} = \text{diag}(\boldsymbol{\sigma}_b)\, \mathbf{R}\, \text{diag}(\boldsymbol{\sigma}_b)$. Gamma priors are assigned to the standard deviations $\boldsymbol{\sigma}_b$, while the correlation matrix $\mathbf{R}$ is given an LKJ prior (Lewandowski et al., 2009). A random sample from the joint posterior in (2.8) is obtained using a Metropolis-within-Gibbs algorithm, with Gibbs steps for the variance parameter and Metropolis steps for the other parameters.

As already mentioned, the roughness penalty from the frequentist P-spline approach (Eilers and Marx, 1996) takes the form of a Gaussian Markov field prior for the B-spline parameters in the Bayesian framework (Rue and Held, 2005),

$$p(\boldsymbol{\gamma}_v \mid \tau_v) \propto \tau_v^{\rho(\mathbf{K})/2} \exp\left(-\frac{\tau_v}{2} \boldsymbol{\gamma}_v^\top \mathbf{K} \boldsymbol{\gamma}_v \right) \;;\; \tau_v \sim \text{Gamma}(\mathfrak{a}_v, \mathfrak{b}_v)$$

$$p(\boldsymbol{\gamma}_{h_0} \mid \tau_{h_0}) \propto \tau_{h_0}^{\rho(\mathbf{K}_0)/2} \exp\left(-\frac{\tau_{h_0}}{2} \boldsymbol{\gamma}_{h_0}^\top \mathbf{K}_0 \boldsymbol{\gamma}_{h_0} \right) \;;\; \tau_{h_0} \sim \text{Gamma}(\mathfrak{a}_{h_0}, \mathfrak{b}_{h_0})$$

where τ_v and τ_{h_0} are smoothing parameters and $\mathbf{K} = \boldsymbol{\Delta}_r^\top \boldsymbol{\Delta}_r$ is the penalty matrix of rank $\rho(\mathbf{K})$ associated to the additive terms f_j, where $\boldsymbol{\Delta}_r$ denotes the rth-order difference penalty matrix (with $r = 2$, say). Similarly, $\mathbf{K}_0$ denotes the penalty matrix associated to the baseline hazard function. The amount of smoothness is controlled by τ_v and τ_{h_0}. It is generally recommended to set $\mathfrak{a}_v$ and $\mathfrak{a}_{h_0}$ equal to 1 and $\mathfrak{b}_v$ and $\mathfrak{b}_{h_0}$ equal to a small number (say 0.005) (Lang and Brezger, 2004). This specific parameterization provides a weakly informative prior that is nearly flat over a wide range of plausible values, allowing the data to effectively determine the optimal level of smoothness while avoiding the numerical instabilities sometimes associated with excessively vague priors. The choice of prior distributions follows standard practice in the literature on joint models and Bayesian P-splines.. Alternatively, a mixture of Gamma distributions could be considered to obtain a robust specification for the roughness penalty prior (Jullion and Lambert, 2007).

We use a second-order difference penalty ($r = 2$) to penalize curvature, such that in the limit of strong smoothing the estimated effect becomes linear. The corresponding difference penalty matrix Δ_2 is given by

$$\Delta_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -2 & 1 \end{pmatrix}.$$

Large values of the smoothing parameters τ_v and τ_{h_0} enforce stronger smoothness and discourage abrupt changes in adjacent spline coefficients, while smaller values allow greater local flexibility.

2.3.3 Identifiability

Because the survival predictor in (2.3) is additive, identifiability constraints are required for the smooth components. Indeed, each smooth term f_j is only identifiable up to an additive constant. Hence, without further restriction, the

decomposition of the predictor into its additive components is not unique. This is the classical identifiability issue encountered in additive models.

Following Wood (2017), we impose sum-to-zero constraints on the smooth functions,

$$\sum_{i=1}^{n} f_j(v_{ij}) = 0, \qquad j = 1, \ldots, J,$$

so that each additive function has zero mean over the observed covariate values. This constraint removes the indeterminacy due to arbitrary additive constants and ensures a unique and interpretable decomposition of the additive predictor.

In practice, the constraint is enforced by centering the B-spline basis matrix $\mathbf{B}_j$. Let $\tilde{\mathbf{B}}_j$ denote the centred basis obtained by subtracting from each column of $\mathbf{B}_j$ its empirical mean. Then $\mathbf{1}^\top \tilde{\mathbf{B}}_j = \mathbf{0}^\top$, which implies that any function represented as $\tilde{\mathbf{B}}_j \boldsymbol{\gamma}_{vj}$ automatically satisfies the sum-to-zero constraint. However, column centering reduces the rank of $\tilde{\mathbf{B}}$ from K to $K - 1$, implying that only $K - 1$ spline coefficients are identifiable. Following the implementation described by Wood (2017), we fix one spline coefficient in each vector $\boldsymbol{\gamma}_{vj}$ to zero and remove the corresponding column from the centred basis matrix $\tilde{\mathbf{B}}$ and from the associated difference penalty matrix $\boldsymbol{\Delta}_r$. This yields a constrained basis representation and the corresponding constrained penalty matrix $\tilde{\mathbf{K}}$.

More details on identifiability issues in additive models and how the constraints solve it can be found in Appendices A2.

2.4 Simulation study

The objective of this section is to evaluate the statistical performance of our approach (FlexibleJM) which allows us to estimate, within a Bayesian framework, a joint model that can contain covariates with nonlinear effects in the survival submodel. We evaluate the finite sample properties of our estimation strategy through a simulation study, focusing on two aspects. First, we assess the ability to estimate smooth additive terms, and second, we compare our results with those obtained with the `JMbayes2` package (Rizopoulos, 2023) assuming linearity for the additive terms. This package enables to fit a joint model under a Bayesian approach but currently does not permit the inclusion of a nonlinear effect of baseline covariates in the survival submodel. It assumes a linear relationship between each survival covariate and the log hazard function. With the `JMbayes2` package, the estimation of the joint

model is performed via the function `jm()`. We name it below the LinearJM approach. The FlexibleJM approach was implemented using parts of the `JMbayes2` package code as a starting point.

2.4.1 Simulation design

In this section, we consider the special case of a Poisson longitudinal response. We also carried out simulations with other distributions for the longitudinal response (Gaussian and binary), the results of which are presented in Appendices A3.1 and A3.2. In Appendix A3.3, we additionally compared our approach with the Bayesian additive joint modeling framework of Köhler et al. (2017), as implemented in the `bamlss` package.

For every setting, we simulate a maximum of 10 repeated measurements for n subjects at a fixed grid of time points $t_p = [0, 2, 4, \ldots, 20]$ based on a true longitudinal model

$$\eta_i(t) = \beta_0 + \beta_1 t_{ij} + \beta_2 x_{1i} + \beta_3 x_{2i} + b_{0i} + b_{1i} t_{ij}$$

where $\beta_0 = 0.55$, $\beta_1 = -0.25$, $\beta_2 = -0.60$, $\beta_3 = 0.30$, $x_{1i} \sim N(0, 0.5^2)$, $x_{2i} \sim \text{Bin}(1, 0.5)$, $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$ with $\text{var}(b_{0i}) = 1$, $\text{var}(b_{1i}) = 0.2^2$ and $\text{cov}(b_{0i}, b_{1i}) = 0.04$. We generate the observed longitudinal measurements from $y_{ij}|\mathbf{b}_i \sim \text{Pois}(\lambda_{ij})$ where $\eta_{ij} = \log(\lambda_{ij})$ and $\eta_{ij} = \eta_i(t_{ij})$. The hazard $h_i(t)$ for every subject is defined as follows:

$$h_i(t \mid \mathcal{N}_i(t), \mathbf{w}_i, \mathbf{v}_i) = h_0(t) \exp\left\{\gamma_1 w_{1i} + \gamma_2 w_{2i} + \sum_{j=1}^{J} f_j(v_{ij}) + \alpha \eta_i(t)\right\}$$

where $\gamma_1 = 0.90$, $\gamma_2 = 0.70$, $\alpha = -0.60$, $w_{1i} \sim N(0, 0.5^2)$, $w_{2i} \sim \text{Bin}(1, 0.5)$ and using a smooth non-linear baseline hazard function $h_0(t)$. For the additive terms part $\sum_{j=1}^{J} f_j(v_{ij})$, we investigated the two following scenarios.

Scenario I. One nonlinear effect:

$$h_i(t) = h_0(t) \exp\left\{\gamma_1 w_{1i} + \gamma_2 w_{2i} + f_1(v_{i1}) + \alpha \eta_i(t)\right\}$$

Scenario II. One nonlinear effect and one linear effect:

$$h_i(t) = h_0(t) \exp\left\{\gamma_1 w_{1i} + \gamma_2 w_{2i} + f_1(v_{i1}) + f_2(v_{i2}) + \alpha \eta_i(t)\right\}$$

where in both cases

$$f_1(v_{i1}) = -\sin\big((v_{i1} + 30)/20\big) - 4\ ;\ f_2(v_{i2}) = -0.07\, v_{i2} + 1.6$$

and the variable v_{i1} is simulated from a Uniform distribution on $[18, 89]$ and v_{i2} is simulated from a Uniform distribution on $[18, 35]$.

Based on $h_i(t)$, simulated survival times are generated for every subject using the approach described in Crowther and Lambert (2013a). The cumulative hazard is evaluated using numerical quadrature and survival times are generated using an iterative algorithm which uses the numerical root finding method of Brent (1973). Every subject is censored after $\max(t_p)$ and we additionally right-censored the time-to-event data by generating subject-specific right-censoring times from a Uniform distribution on $\mathrm{U}(k \times \max(t_p), 1.1 \times \max(t_p))$ to the survival times. The value of k is selected in order to obtain 10% and 30% right-censoring rates (cr).

To ensure convergence, we run the model estimation with 20,000 iterations after an adaptive phase of 3,000 iterations and a burn-in of 3,000. The chains are thinned by 10 in order to finally keep 2000 iterations for each parameter. Geweke statistics (Geweke, 1991) are used as a diagnostic tool for chain convergence. For each regression parameter, we assess bias, the root mean-squared error (RMSE) and the effective coverage of 95% credible intervals (COV). We also assess the integrated root mean-squared error (RMISE) for the estimated additive terms. In summary, we have

$$\begin{aligned}
\mathrm{Bias}_{\hat{\beta}_p} &= \frac{1}{S}\sum_{s=1}^{S}\left(\hat{\beta}_p^{(s)} - \beta_p\right) \\
\mathrm{RMSE}_{\hat{\beta}_p} &= \left\{\frac{1}{S}\sum_{s=1}^{S}\left(\hat{\beta}_p^{(s)} - \beta_p\right)^2\right\}^{1/2} \\
\mathrm{COV}_{\hat{\beta}_p} &= \frac{1}{S}\sum_{s=1}^{S}\mathbb{1}\left(\beta_p \in \mathrm{CI}_{\beta_p}^{(s)}\right) \\
\mathrm{RMISE}_{\hat{f}_j} &= \left\{\frac{1}{S}\sum_{s=1}^{S}\int\left(\hat{f}_j^{(s)}(v) - f_j(v)\right)^2 dv\right\}^{1/2}
\end{aligned}$$

where S is the number of replications, $\hat{\beta}_p$ is the estimated posterior mean of a regression parameter β_p, and $\mathbb{1}$ is an indicator function equal to one if the estimated 95% credible interval (CI) contains the true parameter value. The 95% credible interval is defined by the 2.5% and 97.5% empirical quantiles of the posterior samples, so that it contains the central 95% of the posterior distribution of the parameter. Bias and RMSE remain standard performance metrics in simulation studies, even in a Bayesian framework, as they assess the frequentist properties of Bayesian estimators across repeated samples.

The first simulation is based on Scenario I. We aim to assess the impact of the sample size and of the right-censoring rate on the quality of the model estimation and, more specifically, on the estimation of the nonlinear additive term f_1. We therefore consider three data settings: Setting a where $n = 500$ and cr = 10%, Setting b where $n = 500$ and cr = 30% and Setting c where $n = 250$ and cr = 10%. The second simulation is based on Scenario II which contains two additive terms f_1 and f_2. We consider $n = 500$ and cr = 10% and we compare the estimation obtained with the methods FlexibleJM and LinearJM. We would like to highlight that accounting for the possible nonlinear effects of some survival covariates improves the performance of parameter estimation in the joint model. Simulations are performed using $S = 500$ replicates for Scenario I and II and we used 8 B-spline basis functions to estimate the additive terms.

2.4.2 Simulation results

Figure 2.1 shows the estimation of the nonlinear additive function f_1 for the three data settings under Scenario I. For each graph, the 500 grey curves represent the estimations of the corresponding unknown smooth function which is represented by the black curve. For each data setting, the observed estimates are close to the target, and as expected the more information we have, the less variability there is. Indeed, if we compare Setting a with Setting b, we notice that the estimated curves are closer to the target curve in the first case. This is due to the fact that on average 450 events occur in Setting a versus 350 in Setting b. We also note that for smaller data sets (Setting c), there is more uncertainty in the estimation, especially for values at the extremities of the covariate range. The integrated root mean square error is calculated and compared for each setting. We see in Table 2.2 that the RMISE values for Settings b and c are respectively 1.6 times and 1.9 times larger than those for Setting a. This confirms our previous observation that the estimates obtained in Setting a tend to be more precise than in the other two settings.

The second simulation compares the estimation results obtained with the FlexibleJM and LinearJM approaches within the framework of a joint model allowing for nonlinear covariate effects in the survival submodel. Figure 2.2 displays the true and estimated additive terms f_1 and f_2 for $S = 500$ replications, obtained with the FlexibleJM (top) and LinearJM (bottom) approaches, respectively. The grey dashed curves indicate the pointwise mean of the 500 estimated curves. The average computation times per dataset are similar for FlexibleJM and LinearJM, with values of 1.83 and 1.66 minutes, respectively, using a single core of a 2.6 GHz Intel Core Processor i5-1145G7.

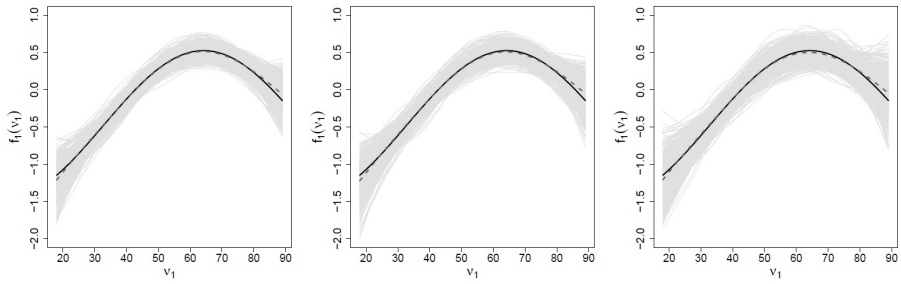

Figure 2.1: True (black) and estimated (grey) smooth additive term f_1 in Settings *a, b, c* (respectively : left, middle, right) under Scenario I. The grey dashed curves indicate the pointwise mean of the estimated curves.

Table 2.2: Integrated root mean square error (RMISE) for the smooth additive terms f_1 and f_2 under Scenario I and Scenario II, obtained with the FlexibleJM and LinearJM approaches. For Scenario I, results are reported for different sample sizes (n) and censoring rates (cr).

Scenario I				
	n	cr	f	FlexibleJM
Setting a :	500	10%	$\hat{f}_1$	0.923
Setting b :	500	30%	$\hat{f}_1$	1.137
Setting c :	250	10%	$\hat{f}_1$	1.778

Scenario II		
f	FlexibleJM	LinearJM
$\hat{f}_1$	0.930	7.881
$\hat{f}_2$	0.105	0.091

In Figure 2.2 (top), the estimated curves closely match the true functions f_1 and f_2, with near-perfect overlap between the dashed and black curves for both effects. This indicates that the estimation of the additive terms using P-splines within the joint modelling framework performs well under the proposed FlexibleJM approach. In contrast, Figure 2.2 (bottom) shows the estimated effects of v_1 and v_2 obtained using LinearJM across the 500 replications. As expected, this approach fails to capture the nonlinear effects of the survival covariates due to model misspecification, since linear effects are assumed throughout.

Table 2.3 compares the performance measures obtained for each regression parameter with each of the two methods. While both approaches show similar performances for the linear regression parameters $\beta_0, \ldots, \beta_3$ in the longitudinal submodel, the FlexibleJM method outperforms LinearJM in estimating the regression parameters γ_1 and γ_2 and the association parameter α, due to the presence of nonlinear effects in the survival part. FlexibleJM achieves lower bias and RMSE and better (frequentist) coverages for the credible intervals in the estimation of the survival regression parameters compared to LinearJM. This is confirmed by the boxplots of the estimated values in Figure 2.3. We also note that the RMISE values for $\hat{f}_2$ are quite similar for both methods (see Table 2.2). This confirms that, compared to the LinearJM method, the precision of the estimates provided by FlexibleJM is comparable for linear effects and is improved for nonlinear effects.

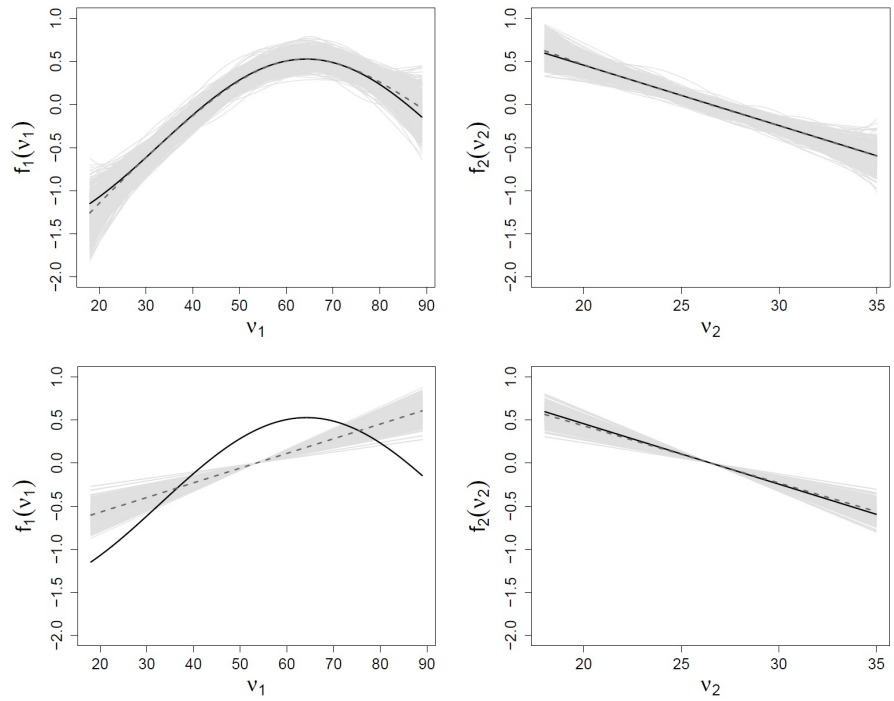

Figure 2.2: True (black) and estimated (grey) smooth additive terms f_1 and f_2 for $S = 500$ dataset replications in Scenario II obtained with the FlexibleJM (top) and LinearJM (bottom) approaches. The grey dashed curves correspond to the pointwise mean of the 500 estimated curves.

In conclusion, our simulations show that the FlexibleJM approach is able to take into account nonlinear effects of survival covariates and provides accurate estimates of the additive terms. Using FlexibleJM, parameter estimation in the joint model outperforms that obtained with LinearJM for the survival parameters.

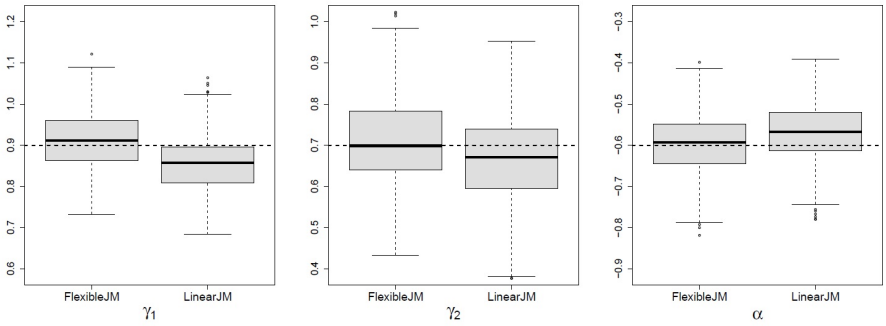

Figure 2.3: Boxplots of the estimated survival regression parameters from S=500 replications in Scenario II comparing methods FlexibleJM and LinearJM. The dashed horizontal lines indicate the true value of the parameters.

Table 2.3: Bias, RMSE and effective coverage of 95% credible intervals for regression parameters for Scenario II under the FlexibleJM and LinearJM approaches.

		FlexibleJM			**LinearJM**		
	True	Bias	RMSE	COV	Bias	RMSE	COV
β_0	0.55	0.002	0.071	0.958	0.002	0.071	0.958
β_1	-0.25	-0.003	0.012	0.944	-0.003	0.011	0.938
β_2	-0.60	-0.002	0.046	0.950	-0.002	0.046	0.948
β_3	0.30	0.002	0.089	0.946	0.001	0.089	0.948
γ_1	0.90	0.010	0.069	0.956	-0.045	0.081	0.856
γ_2	0.70	0.003	0.107	0.964	-0.033	0.115	0.918
α	-0.60	0.007	0.072	0.976	0.030	0.079	0.914

2.5 Application

As introduced in Section 2.1, glioblastoma is a highly malignant form of cancer that develops in the brain. It is a prevalent and extremely aggressive cancer with a very poor prognosis for survival and no curative treatment (Wick et al., 2017; Taal et al., 2014; Tan et al., 2020). The assessment of prognostic factors is an important issue to guide treatment decisions and patient management. Previous research has shown that O6-methylguanine-DNA methyltransferase (MGMT) promoter methylation status, corticosteroid administration at baseline, World Health Organization (WHO) performance status (PS), tumour size and, to some extent age have a significant association with patient survival (Gorlia et al., 2012). All of these factors were measured at baseline. Between 2011 and 2015, a phase III clinical trial was conducted by the European Organisation for Research and Treatment of Cancer (EORTC-26101 trial) (ClinicalTrials.gov, 2021) to determine whether the combined use of lomustine and bevacizumab could enhance overall survival (OS) in patients with first progression of glioblastoma as compared to lomustine alone. In this study, WHO-PS and tumour size were assessed longitudinally throughout the follow-up period. We would therefore like to assess their effect on overall survival over time instead of just accounting for their baseline value at the start of the trial. In addition, corticosteroids are often used to treat patients with glioblastoma to help manage symptoms, but their use must be carefully monitored and managed as their potential side effects can be severe. We therefore want to improve the modelling of the effect of corticosteroids on the risk of death. Given these circumstances, we first want to assess the association between changes over time in WHO-PS and the risk of death in patients with progressive glioblastoma. In a second step, we also want to investigate whether there is an association between changes over time in tumour size and the risk of death. For both analyses, we consider a potential nonlinear effect of the baseline dose of corticosteroids on the risk of death. The dataset analysed contains $n = 516$ patients enrolled across multiple institutions within Europe. Follow-up times varied between 8 and 919 days and 95 (18.4%) patients were right-censored. The median survival time is equal to 267 (95% CI : 249-290). There are 1809 repeated measurements for each biomarker. The mean number of repeated measurements per patient is 3.5 (range: 1 - 15).

2.5.1 Estimation and model assessment

To estimate the joint models presented in Sections 2.5.2 and 2.5.3 in the Bayesian framework, MCMC methods were used with chains of length 20,000 following an adaptive phase of 3,000 iterations and a burn-in of 3,000, see Section

2.3 for methodological details. The trace plots obtained showed no apparent trend, which can be taken as evidence of convergence. Furthermore, we used the Gelman-Rubin diagnostic to confirm that the scale reduction $\hat{R}$ of all parameters are smaller than 1.1.

To compare the competing joint models, we rely on Bayesian model assessment criteria that balance goodness-of-fit and predictive ability. In particular, we report the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002), the Log-Pseudo Marginal Likelihood (LPML) (Ibrahim et al., 2001), and the Widely Applicable Information Criterion, also known as the Watanabe–Akaike Information Criterion (WAIC) (Watanabe, 2010). The DIC is based on the deviance and provides a measure that balances model fit and complexity through the effective number of parameters. The LPML is based on conditional predictive ordinates (Geisser and Eddy, 1979) and provides an assessment of out-of-sample predictive performance from a leave-one-out cross-validation perspective. Finally, the WAIC is a fully Bayesian predictive criterion based on the pointwise log posterior predictive density and can be viewed as an alternative to DIC for complex models. These criteria are widely used in Bayesian model comparison to assess competing models with different levels of complexity. Lower values of DIC and WAIC indicate a better trade-off between fit and complexity, whereas higher values of LPML indicate better predictive performance. We note that these criteria should not be interpreted in absolute terms but rather within a comparative framework, where differences between competing models are used to guide model selection.

To complement these numerical criteria, we also performed a graphical posterior predictive check for the fitted joint models. In the Bayesian framework, posterior predictive checks assess model adequacy by comparing the observed data with replicated datasets generated from the posterior predictive distribution under the fitted model. If the model provides an adequate description of the data, the observed and replicated datasets should display similar distributional features. Figure 2.4 presents histograms of the observed longitudinal responses together with several replicated datasets sampled from the posterior predictive distribution. Overall, no systematic discrepancies are observed between the observed and replicated data, suggesting that the models adequately capture the main distributional features of the longitudinal component of the joint model.

2.5.2 Longitudinal evolution of the WHO performance status

Consider first the evolution over time of the patients' performance status, assessed longitudinally during the follow-up. The WHO Performance Sta-

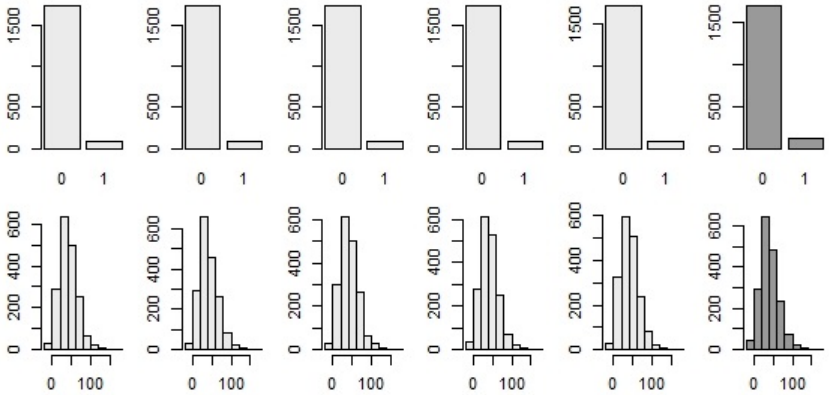

Figure 2.4: Histograms of the observed longitudinal responses (dark grey) and five posterior predictive replications (light grey) for the two longitudinal outcomes.

tus describes a patient's ability to take care of themselves, as well as their daily activity and physical abilities (ECOG-ACRIN Cancer Research Group, 2022). This score runs from 0 to 5, with 0 denoting perfect health and 5 death. The interest is to assess longitudinally the patient's level of capacity (good or poor), knowing that a good performance status is typically characterized by a score in 0-2 (Chahal et al., 2022). The performance status is evaluated at baseline and every 6 weeks until week 24, then every 3 months during the follow-up. The binary indicators $y(t)$ of a poor performance status at time t have a Bernoulli distribution with success probability $\pi_i(t)$. We considered the following joint model,

$$\begin{cases} h_i(t) = h_0(t)\exp\{\gamma_1 w_{\text{LesionB}} + \gamma_2 w_{\text{Age}} + \gamma_3 w_{\text{MGMT}_1} + \gamma_4 w_{\text{MGMT}_2} \\ \qquad\qquad + \gamma_5 w_{\text{MGMT}_3} + f(\text{logDoseC}) + \alpha\eta_i(t)\}, \\ \eta_i(t) = \text{logit}(\pi_i(t)) = \beta_0 + \beta_1 t + b_{0i} + b_{1i} t \end{cases}$$

In the survival submodel, $f(\text{logDoseC})$ is an unknown smooth function estimated in FlexibleJM, but forced to be linear in LinearJM. Covariate LesionB denotes the lesion size at baseline measured by the maximum diameter (in mm) of the largest lesion. Covariate Age denotes the age of the patient at baseline. The categorical covariate MGMT denotes the methylation status of the MGMT promoter with categories 0=unmethylated (Reference category), 1=methylated, 2=undetermined and 3=not done. This last category refers to patients for whom there was no tumour tissue sample of sufficient quality to be able to carry out the test. Finally, logDoseC is constructed using logDoseC = log(DoseC + 1) where DoseC refers to the corticosteroid dose (in mg) administered at baseline. Patients characteristics are described in Ta-

ble 2.4.

Table 2.4: Descriptive statistics of survival covariates and longitudinal biomarkers

Covariates	
Sample size	516
LesionB	
Median	45.0
Std. Deviation	21.1
Range	11.0 - 132.0
Age	
Median	57.8
Std. Deviation	10.8
Range	19.3 - 82.3
logDoseC	
Median	0.4
Std. Deviation	1.5
Range	0.0 - 5.3
MGMT	
0	169 (32.7%)
1	135 (26.2%)
2	96 (18.6%)
3	116 (22.5%)
WHOB	
Poor	488 (94.6%)
Good	28 (5.4%)
Biomarkers	
Number of repeated measurements	1809
Longitudinal values of Lesion	
Median	38.0
sd	23.9
Range	0.0 - 179.0
Longitudinal values of WHO	
Good	1686 (93.2%)
Poor	123 (6.8%)

The estimated parameters of the survival submodel are presented in Table 2.5. The association between changes in WHO-PS values and the risk of death is significantly positive. Poor performance status is significantly associated with a higher risk of death. The hazard ratio for death with methylated

MGMT status as compared with unmethylated MGMT status is $\exp(-0.9015) = 0.41$, which means that the MGMT promoter methylation is an important prognostic factor. The Bayesian model assessment criteria support the FlexibleJM approach, with lower DIC and WAIC values and a higher LPML. While the difference in DIC is relatively small, the larger improvements in WAIC and LPML indicate a better predictive performance of the FlexibleJM model.

Moreover, the left panel of Figure 2.5 shows that our FlexibleJM approach suggests a significant nonlinear effect of the logarithm of the baseline corticosteroid dose logDoseC on the risk of death, conditionally on LesionB, Age, MGMT and the current WHO performance status. Observed values of logDoseC are indicated by symbols, with $\star$ and $\times$ corresponding to dexamethasone and methylprednisolone, respectively, while the symbol + denotes other types of corticosteroids. As presented in Section 2.2.3, we use the approach of Wood (2013) to confirm the significance of each smooth term.

2.5.3 Longitudinal evolution of lesion size

Consider now the association between the longitudinal evolution of the lesion size and the risk of death. The size of the lesion is assessed for the first time at baseline and every 6 weeks until week 24, then every 3 months (Wick et al., 2017). As can be seen from Figure 2.6, part of the patients showed nonlinear evolutions in their lesion size values. To capture the nonlinear subject-specific evolutions, we assume a flexible linear mixed-effects model. We opted for natural cubic splines because the `JMBayes2` package allows the use of the function `ns()` from the package `splines()` in the specification of the longitudinal submodel (Rizopoulos, 2016, 2023). Therefore, we considered the following joint model,

$$\begin{cases} h_i(t) = h_0(t) \exp\{\gamma_1 w_{\text{WHOB}_{\text{Poor}}} + \gamma_2 w_{\text{Age}} + \gamma_3 w_{\text{MGMT}_1} + \gamma_4 w_{\text{MGMT}_2} \\ \qquad\qquad + \gamma_5 w_{\text{MGMT}_3} + f(\text{logDoseC}) + \alpha \eta_i(t)\}, \\ \eta_i(t) = \beta_0 + s(t) + b_{0i} + b_{1i} t \end{cases}$$

where $s(.)$ is a nonlinear function of t, b_{0i} is a random intercept and $b_{1i}t$ is a random slope (Rizopoulos and Ghosh, 2011). The longitudinal measurements of the lesion size $y_i(t)$ are supposed normally distributed conditionally on the random effects, so we have $\eta_i(t) = \mathbb{E}(y_i(t) \mid \mathbf{b}_i)$ with the identity link. The binary covariate WHOB represents the patient's level of capacity (good or poor) at baseline. The model comparison criteria again favour the FlexibleJM approach. Although differences in DIC are modest, the improvements in WAIC

Table 2.5: Estimation results of the survival submodel parameters when we consider the binary version of the WHO-PS as the longitudinal biomarker (WHOL): estimates and 95% credible intervals are presented.

	FlexibleJM		**LinearJM**	
	Est.	CI 95%	Est.	CI 95%
γ_1 (LesionB)	0.3179	[0.1721; 0.4683]	0.3227	[0.1511; 0.5174]
γ_2 (Age)	-0.0087	[-0.1772; 0.1600]	-0.0122	[-0.1829; 0.1449]
γ_3 (MGMT_1)	-0.9015	[-1.3565;-0.5447]	-0.9646	[-1.5594;-0.4977]
γ_4 (MGMT_2)	-0.3690	[-0.8491; 0.1120]	-0.3706	[-0.9193; 0.1116]
γ_5 (MGMT_3)	-0.4076	[-0.7578;-0.1043]	-0.4894	[-0.9265;-0.0587]
α ($\text{Assoc} : \text{WHOL}_{\text{Poor}}$)	0.7137	[0.5097; 0.9887]	0.7204	[0.4689; 1.0991]
f (logDoseC)	(see Figure 2.5.a)		0.1345	[0.0296; 0.2466]
DIC	8443.4		8445.5	
LPML	-4388.2		-4418.8	
WAIC	8775.9		8835.1	

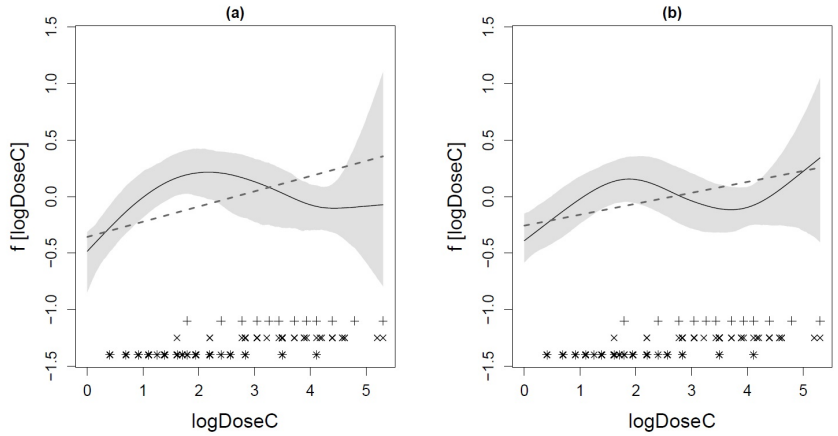

Figure 2.5: Estimated effect of logDoseC **on the log hazard function. Black curves correspond to nonlinear estimates obtained with FlexibleJM, with grey shaded areas representing 95% pointwise credible envelopes. Dotted grey curves show linear estimates obtained with LinearJM. Panel (a) corresponds to the joint model including the longitudinal evolution of binary WHO-PS, whereas panel (b) considers the longitudinal evolution of lesion size.**

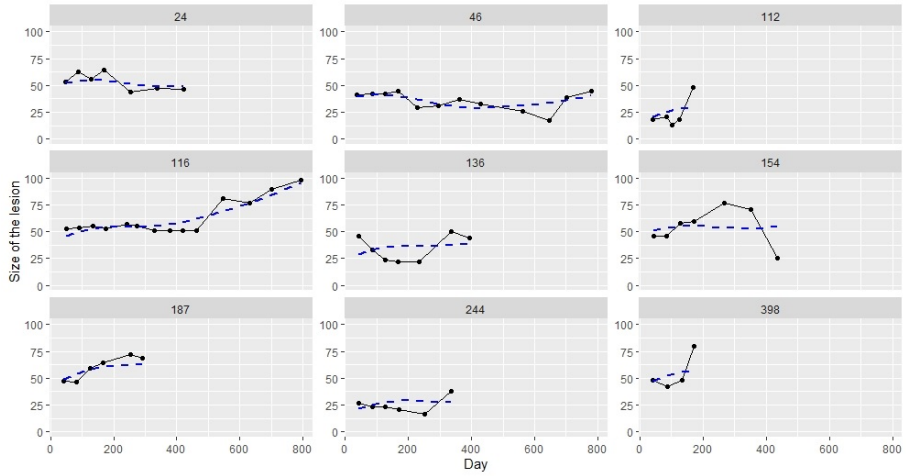

Figure 2.6: Subject-specific longitudinal profiles (black) and model-based predictions (blue) of lesion size for nine randomly selected patients.

and LPML suggest a better predictive ability of the more flexible specification.

The results in Table 2.6 show that the association between the time-varying lesion size and the risk of death is significantly positive. We note that the effect of $MGMT_2$ becomes non significant when the model is estimated with FlexibleJM. As we see in the right panel of Figure 2.5, we still get a significant nonlinear effect of logDoseC at baseline.

The estimated nonlinear effects for the log dose of corticosteroid reported in Figure 2.5 deserve some more comments. Indeed, the two panels in Figure 2.5 show that an increase in logDoseC from 0 mg to 2 mg is associated to an increased risk of death. Thereafter, the risk seems to decrease, and above a logDoseC of 4 mg, the risk stabilises (left panel) or increases again (right panel) depending on the considered longitudinal biomarker. In order to explain this nonlinear behaviour, we conducted a post-hoc analysis and we found that the different types of corticosteroid administered to patients across centres were not directly comparable. In particular, if we compare the two most commonly administered corticosteroid in this study, a 0.75 mg dose of dexamethasone is equivalent to a 4 mg dose of methylprednisolone (Singer and Webb, 2005). The nonlinear estimation of the effect of logDoseC therefore highlighted the non-equivalence of the different corticosteroids used, a feature that the LinearJM approach would not have revealed.

2.6 Discussion

We have presented an extension of joint models for longitudinal and survival data that allows us to account for the nonlinearity of the effects of certain covariates in the survival submodel and the possible non-normality of the longitudinal data. These nonlinear effects are modelled using P-splines and implemented in a Bayesian framework with an automatic tuning of the smoothness of the additive terms. Since the longitudinal process is modelled using GLMMs, it allows for the inclusion of a variety of response variables within the exponential family.

The proposed estimation strategy performs well in different simulation scenarios. Our FlexibleJM approach can effectively accommodate nonlinear effects of covariates in the survival submodel and produce accurate estimates of the additive terms, thereby extending the LinearJM framework, which requires the user to assume linearity or to categorise continuous covariates. Taking these nonlinear effects into account also improves the estimation of the conditional effects of the other covariates. Furthermore, we have shown that even when the data are generated with a linear effect for the survival co-

Table 2.6: Estimation results of the survival submodel parameters when we take the lesion size as longitudinal biomarker (LesionL): estimates and 95% credible intervals (CI) are presented.

	FlexibleJM		**LinearJM**	
	Est.	CI 95%	Est.	CI 95%
γ_1 ($\text{WHOB}_{\text{Poor}}$)	1.2606	[0.7708; 1.7135]	1.2424	[0.7047; 1.7436]
γ_2 (Age)	0.0805	[-0.0224; 0.1821]	0.0704	[-0.0352; 0.1777]
γ_3 (MGMT_1)	-0.5121	[-0.7929;-0.2353]	-0.5374	[-0.8111;-0.2606]
γ_4 (MGMT_2)	-0.1958	[-0.4986; 0.1008]	-0.2306	[-0.5264;-0.0613]
γ_5 (MGMT_3)	-0.1882	[-0.4649;-0.0859]	-0.2486	[-0.5237;-0.0210]
α (Assoc : LesionL)	0.5766	[0.4855; 0.6706]	0.5638	[0.4707; 0.6588]
f (logDoseC)	(see Figure 2.5.b)		0.0967	[-0.0159; 0.2028]
DIC	8743.4		8748.4	
LPML	-4380.1		-4413.3	
WAIC	8759.8		8825.2	

variate, our approach performs almost as well as the classical linear method in estimating this linear effect. These results suggest that the proposed joint model is worth considering for joint modelling of longitudinal and survival data when a nonlinear effect of a survival covariate is suspected.

Finally, we compared our approach with a flexible Bayesian additive joint modeling framework, namely the BAMLSS approach of Köhler et al. (2017) (see Appendix A3.3). The results show that both methods achieve comparable estimation accuracy when incorporating nonlinear effects in the survival submodel. However, our FLEXIBLEJM approach is substantially more computationally efficient, requiring only a few minutes compared to nearly one hour for BAMLSS. In addition, while the BAMLSS framework is currently restricted to Gaussian longitudinal outcomes, our method accommodates a broader class of responses through the use of generalized linear mixed models.

Our analysis of real data showed, across two settings, the importance of accounting for nonlinear effects in the survival submodel. Allowing for a flexible nonlinear effect of the baseline corticosteroid dose led to an improved model fit and predictive performance compared to a purely linear specification. This improvement was supported by Bayesian model assessment criteria, including LPML and WAIC.

An interesting extension of this work would be to incorporate formal statistical tests to assess both the significance of the flexible effects and the presence of nonlinearity. For instance, spline-based approaches such as the `pspline()` formulation proposed by Therneau (2026) allow the decomposition of a covariate effect into a linear component and a nonlinear deviation, enabling separate hypothesis tests for the overall effect and for departures from linearity.

One limitation of our method is that the estimated non-linear effects are common to all patients. FLEXIBLEJM does not allow different nonlinear effects to be computed for different groups of patients.

Within the framework of joint models, several additional extensions are possible. In this chapter, we have used only one longitudinal outcome, but models that with multiple longitudinal outcomes could be desirable. Different types of longitudinal responses could be accommodated by postulating a multivariate generalized linear mixed effects model. This extension is in theory straightforward as long as we assume that the random effects terms take into account all dependencies between the observed data. That is, given random effects, both the longitudinal outcomes as well as the repeated measurements

for each outcome are conditionally independent. Furthermore, the longitudinal outcomes are independent of the time-to-event conditionally on the random effects (Rizopoulos and Ghosh, 2011; Brown et al., 2005). However, the estimation of multivariate joint models can be very complex since the dimension of the variance-covariance matrix for random effects increases with the number of longitudinal outcomes modeled, resulting in a high computational burden (Papageorgiou et al., 2019).

Furthermore, as mentioned in Section 2.2.2, several association structures can be explored to link the longitudinal process with the risk of the event. While the current-value association is the most commonly used specification, alternative formulations such as shared random effects or current slope associations may provide different clinical insights (see Section 1.1.3). The selection of the most appropriate parameterization is an important step in joint model building. Bayesian model assessment criteria such as WAIC and LPML, presented in Section 2.5, provide a useful tool for comparing competing specifications. These criteria help select models that achieve a good balance between model fit and predictive performance. In practice, although these criteria provide a quantitative basis for model selection, the final choice should also take into account biological plausibility and model simplicity, especially when the differences between models are small or practically negligible.

2.7 Appendix A

A1 Preliminary assessment of nonlinear effects

As a preliminary step prior to joint modeling, potential nonlinear effects of continuous covariates can be explored using penalized splines within a Cox proportional hazards model, as implemented by the `pspline()` function in the `survival` package in R (Therneau, 2023a).
In practice, this can be implemented as follows:

```
fit <- coxph(Surv(time, event) ~ pspline(x), data = data)
summary(fit)
termplot(fit)
```

The `pspline()` formulation decomposes the covariate effect into a linear component and a nonlinear deviation. This allows separate hypothesis tests for the overall effect and for departures from linearity, providing a convenient tool for assessing potential nonlinear relationships. The fitted smooth functions can be visualized together with approximate confidence intervals. Details can be found in Therneau (2026). These exploratory analyses can help identify covariates for which nonlinear effects may be relevant in the survival part of the joint model described in Section 2.2.3.

A2 Identifiability in additive models

Additive models provide a flexible framework to capture nonlinear effects of covariates by representing the predictor as a sum of smooth functions. However, this flexibility introduces identifiability issues, since each smooth component is only defined up to an additive constant. As a result, the decomposition of the predictor into individual effects is not unique unless appropriate constraints are imposed. In this section, we illustrate this issue and its resolution using a simple additive model.

Suppose that we have a response variable y and two explanatory variables x_1 and x_2. We define a simple additive model,

$$y_i = \alpha + f_1(x_{i1}) + f_2(x_{i2}) + \epsilon_i,$$

where α is an intercept parameter, the f_j are smooth functions ($j = 1, 2$), and the ϵ_i are independent $N(0, \sigma^2)$ random variables. The additive terms f_j are expressed as

$$f_j(x_{ij}) = \sum_{k=1}^{K} \gamma_{v,jk} B_{jk}(x_{ij}),$$

where $\gamma_{v,jk}$ represents the B-spline coefficient associated with the basis function $B_{jk}(\cdot)$. We define the n-vectors $\mathbf{f}_j = [f_j(x_{1j}), \ldots, f_j(x_{nj})]^\top$, so that $\mathbf{f}_j = \mathbf{B}_j \boldsymbol{\gamma}_j$, where $B_{jk}(x_{ij})$ is the (i,k) element of $\mathbf{B}_j$.

As introduced above, the presence of multiple smooth functions leads to an identifiability problem. The functions f_j are not identifiable from the intercept, nor from each other. Indeed, each smooth function is only estimable up to an additive constant. Any constant can be added to f_1 and subtracted from f_2 without changing the model predictions. For any $x, c \in \mathbb{R}$, we may write

$$\alpha + f(x) = (\alpha + c) + (f(x) - c),$$
$$f_1(x) + f_2(x) = (f_1(x) + c) + (f_2(x) - c).$$

To resolve this identifiability issue, Wood (2017) proposes imposing sum-to-zero constraints:

$$\sum_{i=1}^{n} f_j(x_{ij}) = 0, \qquad \text{equivalently } \mathbf{1}^\top \mathbf{f}_j = 0, \quad j = 1, \ldots, J,$$

where $\mathbf{1}$ is an n-vector of ones. This constraint shifts each function vertically so that its empirical mean is zero.
Since $\mathbf{f}_j = \mathbf{B}_j \boldsymbol{\gamma}_j$, the constraint can be written as $\mathbf{1}^\top \mathbf{B}_j \boldsymbol{\gamma}_j = 0$. This can be enforced by reparameterizing the basis matrix so that $\mathbf{1}^\top \mathbf{B}_j = \mathbf{0}^\top$, ensuring that the constraint holds automatically for any coefficient vector $\boldsymbol{\gamma}_j$. This is achieved by centring the columns of $\mathbf{B}_j$, leading to the centred matrix

$$\tilde{\mathbf{B}}_j = \mathbf{B}_j - \mathbf{1}\mathbf{1}^\top \mathbf{B}_j / n.$$

The corresponding smooth function becomes

$$\tilde{\mathbf{f}}_j = \tilde{\mathbf{B}}_j \boldsymbol{\gamma}_j = \mathbf{B}_j \boldsymbol{\gamma}_j - \mathbf{1}\mathbf{1}^\top \mathbf{B}_j \boldsymbol{\gamma}_j / n = \mathbf{B}_j \boldsymbol{\gamma}_j - \mathbf{1}\bar{f}_j = \mathbf{f}_j - \bar{f}_j,$$

where

$$\bar{f}_j = \frac{1}{n} \mathbf{1}^\top \mathbf{B}_j \boldsymbol{\gamma}_j = \frac{1}{n} \sum_{i=1}^{n} f_j(x_{ij})$$

is the empirical mean of the smooth function. Hence, the constraint simply corresponds to centring each smooth function by subtracting its mean.

To illustrate how the constraint resolves the identifiability issue, consider the simple model $\mathbb{E}(y) = \alpha + f(x)$. Without constraints, for any constant $c \in \mathbb{R}$, we have

$$\mathbb{E}(y) = (\alpha + c) + (f(x) - c),$$

showing that the decomposition is not unique.

Now, impose the centring constraint by defining $\tilde{f}(x) = f(x) - \bar{f}$, where $\bar{f} = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$. The model becomes

$$\mathbb{E}(y) = \alpha + \tilde{f}(x).$$

Consider the same transformation, adding c to the intercept and subtracting it from f. This yields

$$\begin{aligned}\tilde{\mathbb{E}}(y) &= (\alpha + c) + (f(x) - c) - \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - c) \\ &= \alpha + c + f(x) - c - \bar{f} + c \\ &= \alpha + c + \tilde{f}(x),\end{aligned}$$

which differs from $\mathbb{E}(y)$. Hence, the invariance no longer holds, and the model becomes identifiable.

Finally, centring the B-spline matrices implies a rank reduction of $\mathbf{B}_j$ to $K-1$. To ensure that all the spline coefficients can be estimated in a unique way, we fix the Kth element of each spline vector γ_j to zero and delete the Kth column in $\mathbf{B}_j$ and difference matrix $\mathbf{K}$.

A3 Additional simulation settings

This appendix reports additional simulation results for alternative specifications of the longitudinal response, complementing the Poisson case presented in Chapter 1, in order to assess the robustness of the proposed approach to the choice of the longitudinal distribution. In Section A3.1, we consider a Gaussian longitudinal response with $y_{ij}|\mathbf{b}_i \sim \text{N}(\lambda_{ij}, \sigma^2)$ and $\eta_{ij} = \lambda_{ij}$, while in Section A3.2 we consider a binary longitudinal response with $y_{ij}|\mathbf{b}_i \sim \text{Bernoulli}(\lambda_{ij})$ and $\eta_{ij} = \text{logit}(\lambda_{ij})$. In Section A3.3, we consider the same setting as in Section A3.1, but compare our approach with the competing method proposed by Köhler et al. (2017).

A3.1 Gaussian longitudinal response

Table 2.7: Gaussian longitudinal response. Bias, RMSE and coverage for Scenario II under the FlexibleJM and LinearJM approaches.

		FlexibleJM			LinearJM		
	True	Bias	RMSE	COV	Bias	RMSE	COV
β_0	0.55	0.000	0.056	0.958	0.000	0.056	0.958
β_1	-0.25	0.000	0.005	0.954	0.000	0.006	0.954
β_2	-0.60	0.000	0.033	0.958	0.001	0.033	0.960
β_3	0.30	0.000	0.068	0.948	0.000	0.068	0.948
γ_1	0.90	0.003	0.063	0.954	-0.055	0.082	0.788
γ_2	0.70	0.011	0.098	0.964	-0.041	0.107	0.916
α	-0.60	0.006	0.050	0.976	0.042	0.064	0.864

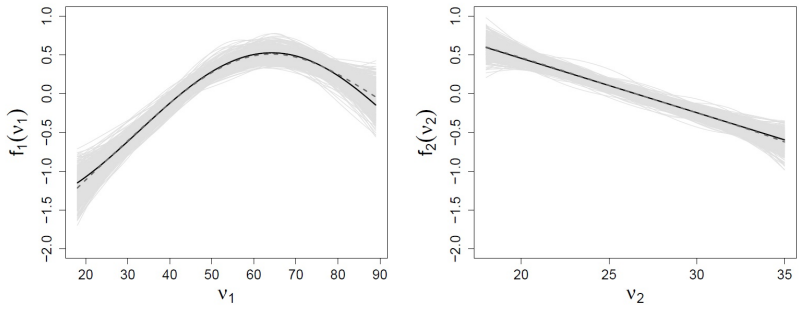

Figure 2.7: Gaussian longitudinal response. True (black) and estimated (grey) smooth additive effects in Scenario II under the FlexibleJM approach.

A3.2 Binary longitudinal response

Table 2.8: Binary longitudinal response. Bias, RMSE and coverage for Scenario II under the FlexibleJM and LinearJM approaches.

		FlexibleJM			LinearJM		
	True	Bias	RMSE	COV	Bias	RMSE	COV
β_0	0.55	0.008	0.105	0.934	-0.009	0.104	0.951
β_1	-0.25	-0.002	0.019	0.929	-0.003	0.030	0.922
β_2	-0.60	-0.004	0.064	0.959	0.003	0.066	0.949
β_3	0.30	0.000	0.115	0.944	0.000	0.117	0.939
γ_1	0.90	0.010	0.078	0.952	-0.037	0.081	0.886
γ_2	0.70	0.012	0.112	0.966	-0.035	0.114	0.931
α	-0.60	-0.005	0.091	0.972	0.019	0.100	0.921

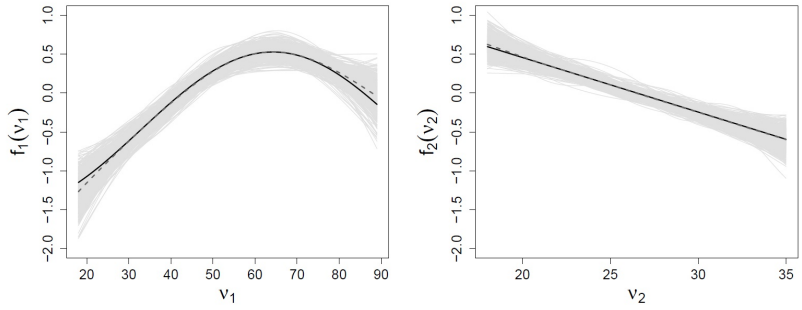

Figure 2.8: Binary longitudinal response : True (black) and estimated (grey) smooth additive effects in Scenario II under the FlexibleJM approach.

A3.3 Comparison with the BAMLSS approach

We compare our proposed method with the flexible Bayesian additive joint modeling approach of Köhler et al. (2017), implemented in the `bamlss` package in R (Umlauf et al., 2018). This framework allows for highly flexible joint models by representing all model components, including longitudinal and survival submodels, using structured additive predictors that may include smooth effects, random effects, and time-dependent components.

The BAMLSS approach employs a derivative-based Metropolis–Hastings algorithm, which constructs proposal distributions using score vectors and Hessian matrices to approximate full conditional distributions (Köhler et al., 2017). This strategy aims to improve estimation stability at the cost of increased computational complexity. For comparison with our FLEXIBLEJM approach, smooth terms are estimated using P-splines with the same number of basis functions in both approaches. In addition, the same number of MCMC iterations is used. The simulated data are identical to those generated under Scenario II described in Section 2.4.1, with a Gaussian longitudinal response. The corresponding results obtained with our FLEXIBLEJM approach are presented in Appendix A3.1.

The results are presented in Table 2.9 and Figure 2.9. The results obtained with the BAMLSS approach are comparable to those obtained with FlexibleJM. The main difference lies in the computational time: the BAMLSS approach takes on average 60.8 minutes, while the FlexibleJM approach requires approximately 2.13 minutes for the same model. This represents a substantial difference in computational efficiency.

Moreover, while the `bamlss` framework allows for various model specifications, the `jm_bamlss` family introduced by Köhler et al. (2017) remains restricted to Gaussian longitudinal submodels. Our methodology overcomes this limitation by integrating generalized linear mixed models, allowing for a wider variety of longitudinal processes while maintaining flexibility in the survival component.

Table 2.9: Gaussian longitudinal response. Bias, RMSE and coverage for Scenario II under the BAMLSS approach.

		BAMLSS		
	True	Bias	RMSE	COV
β_0	0.55	0.015	0.112	0.921
β_1	-0.25	-0.004	0.025	0.926
β_2	-0.60	-0.012	0.078	0.925
β_3	0.30	0.005	0.117	0.930
γ_1	0.90	0.013	0.064	0.925
γ_2	0.70	0.011	0.101	0.938
α	-0.60	-0.018	0.056	0.923

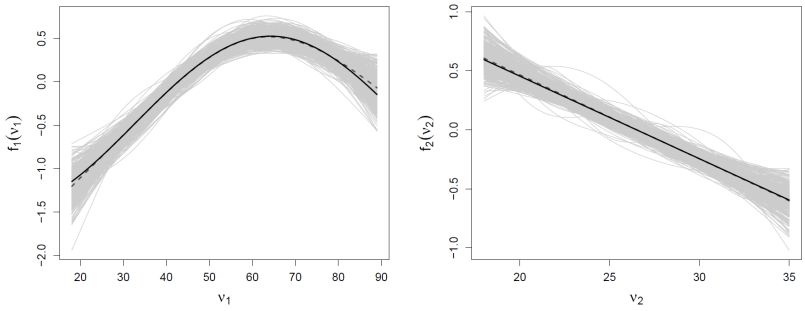

Figure 2.9: Gaussian longitudinal response : True (black) and estimated (grey) smooth additive effects in Scenario II under the BAMLSS approach.

Joint models for HRQoL data with competing dropouts

3

In cancer clinical trials, health-related quality of life (HRQoL) is an important endpoint, providing information on patients' well-being and daily functioning. However, missing data due to premature dropout may lead to biased inference, particularly when dropout is informative. This chapter introduces the JMIRT approach, a novel joint modelling framework for the analysis of multiple longitudinal ordinal outcomes in the presence of informative and competing dropouts. The proposed model links a latent variable underlying HRQoL responses to cause-specific dropout hazards. Unlike classical joint models, which incorporate longitudinal outcomes as covariates in the survival submodel, the JMIRT approach integrates the logarithms of dropout hazards directly into the latent longitudinal process. Through extensive simulation studies, we demonstrate that the proposed method yields robust and unbiased parameter estimates and highlights the importance of accounting for dropout causes. The practical relevance of the approach is illustrated through an application to HRQoL data from patients with progressive glioblastoma.

The material presented in this chapter is based on the paper: Doms H, Lambert P, Legrand C. Joint modelling of longitudinal HRQoL data accounting for the risk of competing dropouts, currently under review at *Journal of the Royal Statistical Society Series C: Applied Statistics*

3.1 Introduction

In cancer clinical trials, overall survival (OS) is the standard endpoint used to demonstrate the clinical benefit of tested treatments. However, while OS provides essential information on treatment efficacy, it may not fully reflect the impact of therapies on patients' daily well-being. To address this limitation, other endpoints have emerged, with health-related quality of life (HRQoL)

becoming a crucial one. HRQoL data are collected using self-administered questionnaires at various time points during patient care and follow-up, providing valuable information on patients' perspectives on their well-being and daily functioning. These data are particularly important for evaluating the overall benefits of new therapies.

Quality of life can be divided into various scales, each measured through one or more items. These items are often measured on a Likert scale, which generates ordinal data. A common approach to analysing these data is the scoring method (Fayers et al., 2001). Item responses are summed to obtain a total score for each scale and each subject at different time points. While this scoring method is widely used, it has significant limitations (Gorter et al., 2015). First, the HRQoL score is often treated as a continuous variable and analysed using linear mixed models (LMM) under the assumption of normality. However, the total score behaves more like an ordinal variable and may show asymmetry due to ceiling or floor effects, violating these assumptions. Furthermore, the gaps between adjacent response categories on Likert scales (e.g., "not at all," "a little," "quite a bit," "very much") are not necessarily uniform, but the scoring method assumes they are. Additionally, subjects with different item responses can end up with the same total score, failing to distinguish between their distinct response patterns, which results in an important loss of information.

Given these challenges, more appropriate methods are needed to account for the ordinal nature of HRQoL data. Item response theory (IRT) models offer an alternative by linking individuals' item responses (raw data) to a latent variable representing the underlying HRQoL component of interest (Van der Linden, 2016). In longitudinal analyses, some studies have extended IRT models by regressing the latent variable on predictors and incorporating subject-specific random effects (Verhagen and Fox, 2013; Wang et al., 2017).

Another important issue with HRQoL data is dropout, which is particularly frequent in cancer clinical trials. Patients may stop completing questionnaires before the end of the clinical trial due to various reasons, such as disease progression or death. When dropout is related to such factors, it is considered informative and should be taken into account in HRQoL analyses to avoid biased results. To address this, joint models (JMs) can be used, as they can model the association between dropout and HRQoL outcomes. Several comprehensive overviews of the joint modelling framework are available (Tsiatis and Davidian, 2004; Rizopoulos, 2012a; Papageorgiou et al., 2019; Wu et al., 2012). Touraine et al. (2023) demonstrated the importance of using joint models for HRQoL data in the presence of informative dropout. Similarly, Cuer

et al. (2022) highlighted the need to distinguish between different causes of dropout by using competing risk joint models. It is indeed crucial to distinguish between informative and non-informative dropouts. Lastly, IRT models have been applied within the joint modelling framework for longitudinal ordinal data. For example, Luo (2014) assessed the impairment caused by Parkinson's disease using a JM with a longitudinal IRT model, but did not account for competing dropout risks.

In this chapter, we aim to develop a model that allows the longitudinal analysis of multiple ordinal categorical outcomes while accounting for different dropout risks. Our primary objective is to improve the modelling of the longitudinal process itself, which remains the main quantity of interest, rather than focusing on the survival component as is commonly done in joint modelling frameworks. We propose a joint modelling framework that uses an IRT model for the multivariate ordinal longitudinal outcomes and a proportional cause-specific hazards model for the dropout risks. Additionally, we aim to improve the analysis of longitudinal data by incorporating survival information directly into the longitudinal submodel. While classical joint models focus on survival information and treat longitudinal data as a covariate in the survival submodel, our approach prioritizes the longitudinal component by incorporating log dropout risks as covariates in the latent longitudinal process. This enables us to account for the potential effects of dropout risks when estimating the longitudinal latent trait.

The rest of the chapter is structured as follows: In Section 3.2, we describe the formulation of our proposed joint model. Section 3.3 presents a Bayesian estimation procedure for the model. In Section 3.4, we discuss a simulation study to assess the statistical performance of the model and Section 3.5 illustrates our approach by analysing responses to the QLQ-C30 QOL questionnaires completed by glioblastoma patients. Finally, Section 3.6 concludes the chapter with a discussion.

3.2 Methodology

3.2.1 Longitudinal submodel

We consider a cohort of n patients followed up over time. For every individual ($i = 1, \ldots, n$), we observe K longitudinal ordinal outcomes measuring the same construct of interest, and each outcome takes values in L_k categories. The longitudinal values are denoted $y_{ijk} = y_{ik}(t_{ij})$ for outcome k ($k = 1 \ldots, K$) observed at time t_{ij} ($j = 1, \ldots, N_i$).

3.2.1.1 Latent process

We define $\eta_{ij} = \eta_i(t_{ij})$ as a latent variable that represents an underlying trait or ability of subject i at occasion j that cannot be directly observed but can be assessed through the longitudinal responses provided to the different items. The trajectory of η over time for patient i is described using a linear mixed model as follows :

$$\begin{cases} \eta_i(t) = \mathbf{x}_i(t)^\top \boldsymbol{\beta} + \mathbf{z}_i(t)^\top \mathbf{b}_i \\ \mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}) \end{cases}$$

Here, $\boldsymbol{\beta}$ denotes the vector of fixed effects associated with a time-dependent design vector $\mathbf{x}_i(t)$, which typically includes an intercept, a time-dependent slope, and can adjust for other covariates. Meanwhile, $\mathbf{b}_i$ represents the vector of random effects linked to $\mathbf{z}_i(t)$. The fixed effects term describes the average trajectory of the latent process at the population level, while the random effects term captures individual-specific deviations from this trajectory.

We assume a multivariate normal distribution for the subject-specific random effects. Although this assumption is standard in joint modelling, concerns may arise regarding potential misspecification. Previous work has shown that parameter estimates and standard errors in shared-parameter joint models are generally robust to moderate deviations from normality, particularly when the number of repeated measurements per subject is adequate (Rizopoulos et al., 2008; Papageorgiou et al., 2019).

3.2.1.2 Graded response model

We now define the link between the latent process η and each longitudinal response. To achieve this, we use Samejima's ordinal response model, also known as the graded response model (GRM) (Samejima, 1969). The model incorporates two key parameters: for each outcome, there is an associated discrimination parameter a_k, and a threshold parameter $\boldsymbol{d}_k = [d_{k,1}, \ldots, d_{k,L_k-1}]$. We define the cumulative probability for the kth ordinal outcome as

$$\begin{aligned} &\text{logit}\{\Pr(y_{ik}(t) \leq l \mid \eta_i(t))\} = d_{kl} - a_k \eta_i(t), \qquad l = 1, \ldots, L_k - 1, \\ \Leftrightarrow \quad &\Pr(y_{ik}(t) \leq l \mid \eta_i(t)) = \frac{1}{1 + \exp\{-(d_{kl} - a_k \eta_i(t))\}}. \end{aligned}$$

Here, $\eta_i(t)$ denotes the latent ability level of subject i at time t, with higher values of $\eta_i(t)$ indicating a lower ability level. Parameter $a_k > 0$ measures the discriminative ability of item k, meaning how well the item can differentiate between patients at different trait levels. Parameter d_{kl} determines, in combination with a_k, the latent trait value $\eta = d_{kl}/a_k$ at which the probability

of responding in category l or lower equals 50%. The threshold parameters satisfy the ordering constraint $d_{k1} < \cdots < d_{k,L_k-1}$. The boundary conditions of the cumulative response probabilities are

$$\Pr(y_{ik}(t) \leq 0 \mid \eta_i(t)) = 0, \qquad \Pr(y_{ik}(t) \leq L_k \mid \eta_i(t)) = 1.$$

The probability of observing category l is then given by, for $l = 1, \ldots, L_k$,

$$\Pr(y_{ik}(t) = l \mid \eta_i(t)) = \Pr(y_{ik}(t) \leq l \mid \eta_i(t)) - \Pr(y_{ik}(t) \leq l-1 \mid \eta_i(t)).$$

The graded response model belongs to the class of polytomous IRT models specifically designed for ordinal responses with more than two ordered categories, as typically encountered in HRQoL questionnaires. Unlike binary IRT models, it explicitly accounts for the ordered nature of Likert-type response scales and allows each item to have its own discrimination parameter, thereby capturing differences in measurement precision across items. This makes the GRM particularly well suited for analysing multi-item scales composed of ordinal responses, such as those found in the QLQ-C30 questionnaire. The validity of this modelling approach nevertheless relies on several standard IRT assumptions, which are discussed in Section 1.2.3.

Identifiability constraints are required because the latent trait is not directly observed. Without such constraints, the model is invariant to translations and rescalings of the latent variable, meaning that different combinations of threshold, discrimination, and latent process parameters can produce identical response probabilities. To ensure identifiability, it is therefore necessary to fix the location and scale of the latent trait. A common approach consists in selecting an anchor item whose parameters are constrained to define a reference origin and unit. This typically involves fixing one discrimination parameter (e.g., $a_1 = 1$) and one threshold (e.g., $d_{1,1} = 0$).

3.2.2 Survival submodel

As introduced in Section 3.1, HRQoL data are often affected by premature termination of questionnaire completion, known in this context as dropout. Since dropout can be either informative (disease related) or non-informative (non-disease related) depending on its cause, it must be accounted for in HRQoL analyses to avoid biased results. To address this, we use a competing risks model for the time-to-dropout, allowing us to distinguish between two types of dropout: dropouts unrelated ($p = 1$) or related ($p = 2$) to a clinical event such as death or disease progression. We assume the latter to be informative, in contrast to the former which is assumed to be uninformative. The event indicator δ_{ip} in $\delta_i = [\delta_{i1}, \delta_{i2}]$ takes value 1 if the i-th individual drops

out due to cause p before administrative censoring, and 0 otherwise (Legrand, 2021).

Let T_i^* be the time at which the dropout occurred for the i-th individual. We observe $T_i = \min(T_i^*, C_i)$, where C_i is the right administrative censoring time (end of study). We model the time-to-dropout using a proportional cause-specific hazards model (Putter et al., 2007). For patient i, the cause-specific hazard of dropout due to cause p is given by

$$h_{ip}(t) = h_{0p}(t) \exp\left\{\boldsymbol{\gamma}_p^\top \mathbf{w}_i + \boldsymbol{\alpha}_p m_p(t, \mathbf{b}_i)\right\}, \qquad t > 0$$

where $h_{0p}(t)$ denotes the cause-specific baseline hazard for cause p ($p = 1, 2$) and $\mathbf{w}_i = [w_{i1}, \ldots, w_{iG}]^\top$ is the vector of baseline survival covariates for subject i with cause-specific regression coefficients $\boldsymbol{\gamma}_p$. Parameters $\boldsymbol{\alpha}_p$ quantify the association between the longitudinal latent process and the survival process. Different association functions $m_p(t, \mathbf{b}_i)$ can be specified. The current-value and the random effects association structures are frequently used in the literature (Rizopoulos, 2012a). These are respectively defined by $m_p(t, \mathbf{b}_i) = \eta_i(t)$ and $m_p(t, \mathbf{b}_i) = \mathbf{b}_i$ where in the latter case $\boldsymbol{\alpha}_p$ is a vector of association parameters of the same dimension as $\mathbf{b}_i$.

In the joint modelling framework, the baseline hazard should not be left unspecified; therefore we define the logarithm of the baseline hazard functions using B-splines (Lambert and Eilers, 2005; Rizopoulos, 2016) :

$$\log\{h_{0p}(t)\} = \sum_{u=1}^{U_p} \gamma_{h_{0p},u} B_u(t)$$

where $\gamma_{h_{0p},u}$ is the regression coefficient associated to the uth function of a large B-spline basis associated to equidistant knots on $(0, \max_i T_i)$.

3.2.3 Accounting for dropout risks to enhance the latent process

Finally, we extend the previous model to refine the modelling of the latent process by incorporating survival information into the longitudinal submodel. Specifically, we aim to account for the potential effects of the different hazards associated with competing dropout risks on the longitudinal latent process. To achieve this, we define and include the following covariate in the latent process model:

$$\mathbf{v}_i(t)^\top = [\log \tilde{h}_{i1}(t), \ldots, \log \tilde{h}_{iP}(t)],$$

where the pseudo-hazards $\tilde{h}_{ip}(t)$ are defined as

$$\tilde{h}_{ip}(t) = h_{0p}(t) \exp\left(\boldsymbol{\gamma}_p^\top \tilde{\mathbf{w}}_i + \boldsymbol{\alpha}_p \mathbf{b}_i\right),$$

and $\tilde{\mathbf{w}}_i$ contains only the covariates that are shared between the survival and longitudinal submodels. This choice ensures structural coherence between the two submodels and avoids introducing additional survival-only covariates into the latent process. If all covariates are shared, then $\tilde{h}_{ip}(t) = h_{ip}(t)$. This extension enables us to capture the influence of these hazards on the trajectory of the latent variable $\eta_i(t)$, and subsequently on the longitudinal responses. This part is crucial to ensure that our model adequately reflects the complexities of the relationship between longitudinal HRQoL data and dropout events. It is also important to note that, since we introduce the logarithm of the hazard functions as covariates to describe the latent process, we have to use the random effects association structure to ensure identifiability in the model. This is crucial because using an alternative structure, such as $m_p\{t, \mathbf{b}_i\} = \eta_i(t)$, would introduce circular dependencies that hinder parameter estimation. To ensure identifiability, we restrict the association structure to depend only on the random intercept, i.e., we set $m_p(t, \mathbf{b}_i) = b_{i0}$. The extended model is therefore defined as:

$$\begin{cases} h_{ip}(t) = h_{0p}(t) \exp\left(\boldsymbol{\gamma}_p^\top \mathbf{w}_i + \alpha_p b_{i0}\right), \\ \eta_i(t) = \mathbf{x}_i(t)^\top \boldsymbol{\beta} + \mathbf{z}_i(t)^\top \mathbf{b}_i + \mathbf{v}_i(t)^\top \boldsymbol{\lambda}, \\ \text{logit}\{\Pr(y_{ik}(t) \leq l \mid \eta_i(t))\} = d_{kl} - a_k \eta_i(t) \end{cases} \tag{3.1}$$

where the parameter vector $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_P]^\top$ characterizes the association between the risk of dropout for the different causes and the latent variable of interest $\eta_i(t)$. An estimated λ_p which is not significantly different from zero indicates that the risk of dropout due to cause p has no statistically significant impact on the latent variable $\eta_i(t)$. Conversely, a significantly positive λ_p implies that a higher risk of dropout for cause p is associated with an increased value of $\eta_i(t)$. As introduced earlier, we consider two competing risks of dropout: one unrelated to the progression of the disease and death ($p = 1$) and another related to the occurrence of disease progression or death ($p = 2$). In this case, we define $\mathbf{v}_i(t)^\top = [\log \tilde{h}_{i1}(t), \log \tilde{h}_{i2}(t)]$.

Although dropout due to cause $p = 1$ is clinically considered unrelated to disease progression, this does not guarantee statistical independence from the longitudinal HRQoL process. This motivates the competing-risks framework introduced in Section 3.2.2 and the incorporation of the logarithm of both cause-specific hazards into the specification of the latent process, allowing the potential association between each dropout mechanism and the underlying HRQoL trajectory to be assessed.

3.3 Bayesian estimation

We apply a Bayesian approach where inference is based on the posterior of the model parameters. In particular, we estimate the parameters of the extended joint model (3.1) using Markov Chain Monte Carlo (MCMC) methods.

3.3.1 Likelihood function

In the framework of joint models for longitudinal and survival data, the likelihood is derived under the common assumption that the longitudinal and survival processes are independent conditionally on the random effects $\mathbf{b}_i$. Moreover, the longitudinal measurements $\mathbf{y}_i$ of each individual are assumed independent given the random effects. We also assume that the individual-specific random effects are independent of the observed covariates.

The expression of the marginal likelihood contribution for the i^{th} subject is written as

$$L_i(\theta) = \int p(T_i, \delta_i \mid \mathbf{b}_i; \theta_s) \left\{ \prod_{j=1}^{N_i} \prod_{k=1}^{K} p\big(y_{ik}(t_{ij}) \mid \mathbf{b}_i; \theta_y\big) \right\} p(\mathbf{b}_i; \theta_b)\, d\mathbf{b}_i. \quad (3.2)$$

where $\theta = (\theta_s, \theta_y, \theta_b)$ contains θ_s the parameters for the survival outcomes, θ_y the parameters for the longitudinal outcome and θ_b the parameters of the random-effect covariance matrix. Function $p(.)$ generically denotes the corresponding probability or density function. The marginal log-likelihood function formulated for all subjects is obtained by $l(\theta) = \sum_{i=1}^{n} \log L_i(\theta)$.

The normality assumption for the distribution of random effects implies that

$$p(\mathbf{b}_i; \theta_b) = (2\pi)^{-q/2} \det(\mathbf{D})^{-1/2} \exp\big(-\mathbf{b}_i^\top \mathbf{D}^{-1} \mathbf{b}_i / 2\big) \quad (3.3)$$

where q denotes the dimension of the random effect vector.

The extended expression for the conditional likelihood contribution of the ith

subject in the survival submodel is given by (Putter et al., 2007)

$$
\begin{aligned}
p(T_i,\delta_i \mid \mathbf{b}_i;\theta_s) &= \prod_{p=1}^{P} \left[h_{ip}(T_i \mid \mathbf{b}_i;\theta_s)\right]^{\delta_{ip}} S_{ip}(T_i \mid \mathbf{b}_i;\theta_s) \\
&= \prod_{p=1}^{P} \left[h_{0p}(T_i)\exp\{\boldsymbol{\gamma}_p^\top \mathbf{w}_i + \alpha_p m_p(T_i,\mathbf{b}_i)\}\right]^{\delta_{ip}} \\
&\quad \times \exp\left(-\int_0^{T_i} h_{0p}(s)\exp\{\boldsymbol{\gamma}_p^\top \mathbf{w}_i + \alpha_p m_p(s,\mathbf{b}_i)\}\,ds\right) \\
&= \prod_{p=1}^{P} \left[\exp\left\{\sum_{u=1}^{U_p} \gamma_{h_{0p},u} B_u(T_i) + \boldsymbol{\gamma}_p^\top \mathbf{w}_i + \alpha_p m_p(T_i,\mathbf{b}_i)\right\}\right]^{\delta_{ip}} \\
&\quad \times \exp\left(-\exp\left\{\boldsymbol{\gamma}_p^\top \mathbf{w}_i\right\}\int_0^{T_i} \exp\left\{\sum_{u=1}^{U_p} \gamma_{h_{0p},u} B_u(s) + \alpha_p m_p(s,\mathbf{b}_i)\right\} ds\right)
\end{aligned}
\tag{3.4}
$$

The probability contribution in Eq. (3.2) for the observed longitudinal value of the kth item response at time t can be expressed as follows:

$$
\begin{aligned}
p\big(y_{ik}(t) \mid \mathbf{b}_i;\theta_y\big) &= \prod_{l=1}^{L_k} \Pr\big(y_{ik}(t)=l \mid \eta_i(t)\big)^{\mathbb{1}\{y_{ik}(t)=l\}} \\
&= \prod_{l=1}^{L_k} \Big[\Pr\big(y_{ik}(t)\le l \mid \eta_i(t)\big) - \Pr\big(y_{ik}(t)\le l-1 \mid \eta_i(t)\big)\Big]^{\mathbb{1}\{y_{ik}(t)=l\}} \\
&= \Big\{\operatorname{expit}\big(d_{k1} - a_k\eta_i(t)\big)\Big\}^{\mathbb{1}\{y_{ik}(t)=1\}} \\
&\quad \times \prod_{l=2}^{L_k-1} \Big\{\operatorname{expit}\big(d_{kl} - a_k\eta_i(t)\big) - \operatorname{expit}\big(d_{k,l-1} - a_k\eta_i(t)\big)\Big\}^{\mathbb{1}\{y_{ik}(t)=l\}} \\
&\quad \times \Big\{1 - \operatorname{expit}\big(d_{k,L_k-1} - a_k\eta_i(t)\big)\Big\}^{\mathbb{1}\{y_{ik}(t)=L_k\}}
\end{aligned}
\tag{3.5}
$$

where $\operatorname{expit}(x) = \frac{1}{1+e^{-x}}$.

A numerical method must be used for the evaluation of the integral in (C1) because it has no closed-form solution. We approximate it using the 15-point Gauss-Kronrod quadrature formula (Rizopoulos, 2016).

3.3.2 Priors and posterior distributions

Bayesian estimation of model parameters is based on the posterior distribution. The expression for the posterior distribution of the model parameters is derived under the assumptions defined in Section 3.3.1. Using Bayes theorem, we obtain

$$p(\theta, \mathbf{b} \mid \mathbf{T}, \delta, \mathbf{y}) \propto \prod_{i=1}^{n} p(T_i, \delta_i | \mathbf{b}_i; \theta_s) \prod_{j=1}^{N_i} \prod_{k=1}^{K} p\big(y_{ik}(t_{ij}) | \mathbf{b}_i; \theta_y\big)\, p(\mathbf{b}_i; \theta_b)\; p(\theta) \tag{3.6}$$

where θ denotes the full parameter vector with joint prior $p(\theta)$. The expressions for $p(T_i, \delta_i \mid \mathbf{b}_i; \theta)$, $p(y_{ik}(t) \mid \mathbf{b}_i; \theta)$ and $p(\mathbf{b}_i; \theta_b)$ can be found in (C1), (3.5) and (C1), respectively.

We use standard prior distributions for θ (Papageorgiou et al., 2019; Alsefri et al., 2020). In particular, independent univariate normal priors are used for $\boldsymbol{\beta}$ (the vector of fixed effects in the longitudinal submodel), for $\boldsymbol{\lambda}$ (the vector of risk-related association parameters), for $\boldsymbol{\gamma}$ (the vector of regression coefficients in the survival submodel) and for $\boldsymbol{\alpha}$ (the vector of random effects association parameters). We let all discrimination parameters a have Uniform$[0, 5]$ as prior distribution. Several strategies can be used to impose order constraints on the threshold parameters. One commonly used approach consists in reparameterizing the thresholds through their successive first-order differences and assigning weakly informative half-normal priors $\mathcal{N}^+(0, 10)$ to these differences, which ensures that the thresholds remain ordered. Another widely used strategy relies on directly constraining the thresholds by imposing appropriate bounds on their prior distributions, where each $d_{k,l}$ is assigned a uniform prior distribution with lower bound $d_{k,l-1}$. For the variance-covariance matrix of the random effects $\mathbf{D}$, we take an Inverse-Wishart. The flexibility of the log baseline hazards expressed as a linear combination of a large number of B-splines can be counterbalanced using a Gaussian Markov field prior for $\boldsymbol{\gamma}_{h_{0p}}$ (Eilers and Marx, 1996; Lambert and Eilers, 2005) :

$$p(\boldsymbol{\gamma}_{h_{0p}} \mid \tau_p) \propto \tau_p^{\rho(\mathbf{K}_p)/2} \exp\Big(-\frac{\tau_p}{2}\, \boldsymbol{\gamma}_{h_{0p}}^{\top} \mathbf{K}_p \boldsymbol{\gamma}_{h_{0p}}\Big)$$

$$\tau_p \sim \text{Gamma}(\mathfrak{a}_p, \mathfrak{b}_p)$$

where τ_p is a roughness penalty parameter and $\mathbf{K}_p = \boldsymbol{\Delta}_r^{\top} \boldsymbol{\Delta}_r$ is the penalty matrix of rank $\rho(\mathbf{K}_p) = U_p - r$ associated with the baseline hazard function h_{0p}, where $\boldsymbol{\Delta}_r$ denotes the rth difference penalty matrix (with $r = 2$, say). The amount of smoothness is controlled by τ_p. It is generally recommended to set $\mathfrak{a}_p$ equal to 1 and $\mathfrak{b}_p$ equal to a small number (say 0.005) (Lang and Brezger,

2004).

A random sample from the joint posterior in (4.2) is obtained using a Metropolis-within-Gibbs algorithm with Gibbs steps for the variance parameter and Metropolis steps for the other parameters. The detailed steps of the algorithm are outlined in Algorithm 1. The covariance matrices of the normal proposal distributions for the random walk Metropolis algorithm are carefully defined using preliminary estimates for each submodel separately. These proposal distributions are further tuned during an adaptive phase consisting of A iterations. Afterward, a burn-in period of B iterations is performed, followed by I additional iterations. Finally, the chains are thinned according to a thinning interval of T. We end up with chains of size I/T for each parameter. The full conditional distributions of selected parameters involved in the Gibbs updates described in Algorithm 1 are provided below.

$$\begin{aligned} \tau_p \mid \boldsymbol{\gamma}_{h_{0p}} &\sim \text{Gamma}\left(\mathfrak{a}_p + \frac{\rho(\mathbf{K}_p)}{2}, \mathfrak{b}_p + \frac{\boldsymbol{\gamma}_{h_{0p}}^\top \mathbf{K}_p \boldsymbol{\gamma}_{h_{0p}}}{2}\right) \\ \mathbf{D} \mid \mathbf{b}_i &\sim \text{Inverse-Wishart}\left(q + n,\ q \times D^0 + \sum_{i=1}^{n} \mathbf{b}_i \mathbf{b}_i^\top\right) \end{aligned} \tag{3.7}$$

where q is the dimension of the random effects $\mathbf{b}_i$ and D^0 denotes the scale matrix of the inverse-Wishart prior assigned to $\mathbf{D}$. The rank of the penalty matrix $\mathbf{K}_p$ is denoted by $\rho(\mathbf{K}_p)$.

3.4 Simulation study

In this section, we present simulation studies with different settings designed to evaluate the performance of our approach, JMIRT, which estimates an extended joint model using a Bayesian framework to analyse ordered categorical data while accounting for the effects of different types of dropouts.

3.4.1 Simulation design

Setting I: We generate 500 datasets, each with a sample size of $n = 500$ subjects and a maximum of 10 visits per subject. We consider three ordinal outcomes ($K = 3$), each with 4 categories ($L_k = 4$ for all k). The maximum follow-up time is set to 10. Data are generated according to the following

Algorithm 1 MCMC estimation of joint model using JMIRT approach

Input: Longitudinal data (y_{ijk}, t_{ij}) for $i = 1, \ldots, n$ individuals, survival data (T_i, δ_i), prior distributions for all parameters.

Initialize:
Set initial values $\boldsymbol{\beta}^0, \boldsymbol{\lambda}^0, \mathbf{b}_i^0, a^0, d^0, \mathbf{D}^0, \boldsymbol{\gamma}_p^0, \boldsymbol{\gamma}_{h_{0p}}^0, \tau_p^0, \alpha_p^0$ for all i and p.

Set number of MCMC iterations I, adaptive phase A, burn-in B, and thinning T. Total number of iterations is $M = A + B + I$.

for $m = 1, \ldots, M$ **do**

Step 1: Update longitudinal submodel parameters.

Update $\boldsymbol{\beta}^m$ using Metropolis-Hastings:

- A proposal $\boldsymbol{\beta}^*$ is drawn from a normal proposal distribution.
- Compute acceptance ratio:

$$R = \frac{\pi(\boldsymbol{\beta}^* \mid \text{other parameters})}{\pi(\boldsymbol{\beta}^{m-1} \mid \text{other parameters})}.$$

- Set $\boldsymbol{\beta}^m = \boldsymbol{\beta}^*$ when $U \sim \text{Uniform}(0,1) < R$. Otherwise, $\boldsymbol{\beta}^m = \boldsymbol{\beta}^{m-1}$.

Update $\boldsymbol{\lambda}^m$ using Metropolis-Hastings.

Update a^m and d^m using Metropolis-Hastings.

Update $\mathbf{b}_i^m$ using Metropolis-Hastings.

Update $\mathbf{D}^m$ using Gibbs steps (see Equation 3.7).

Step 2: Update survival submodel parameters.

for $p = 1, \ldots, P$ **do**

Update $\boldsymbol{\gamma}_p^m, \boldsymbol{\gamma}_{h_{0p}}^m, \alpha_p^m$ using Metropolis-Hastings.

Update τ_p^m using Gibbs steps (see Equation 3.7).

Store $\theta^m = (\boldsymbol{\beta}^m, \boldsymbol{\lambda}^m, \mathbf{b}_i^m, a^m, d^m, \mathbf{D}^m, \boldsymbol{\gamma}_p^m, \boldsymbol{\gamma}_{h_{0p}}^m, \tau_p^m, \alpha_p^m)$ for all i.

Output: Posterior samples $\{\theta^m\}_{m=\text{seq}(A+B+1,M,\text{by}=T)}$.

extended joint model:

$$\begin{cases} h_{i1}(t) = h_{01}(t)\exp\left(\gamma_1 w_i + \alpha_1 b_{0i}\right) \\ h_{i2}(t) = h_{02}(t)\exp\left(\gamma_2 w_i + \alpha_2 b_{0i}\right) \\ \eta_i(t) = \boldsymbol{\beta}_1 t + \boldsymbol{\beta}_2 w_i + b_{0i} + b_{1i} t + \boldsymbol{\lambda}_1 \log h_{i1}(t) + \boldsymbol{\lambda}_2 \log h_{i2}(t) \\ \text{logit}\{\Pr(y_{ik}(t) \leq l \mid \eta_i(t))\} = d_{kl} - a_k \eta_i(t) \end{cases}$$

where $l = 1, \ldots, 4$, $w_i \sim \text{Bin}(1, 0.50)$, $b_{0i} \sim \mathcal{N}(0, 0.75^2)$, $b_{1i} \sim \mathcal{N}(0, 0.25^2)$ and $\text{corr}(b_{0i}, b_{1i}) = 0.4$. To facilitate estimation, we exploit an equivalent reparametrized version of the longitudinal submodel. Noting that the hazards $h_{i1}(t)$ and $h_{i2}(t)$ depend on both w_i and b_{0i}, which also appear in the longitudinal predictor $\eta_i(t)$, we rewrite the latter as:

$$\eta_i(t) = \beta_1 t + \beta_2^* w_i + b_{0i}^* + b_{1i} t + \lambda_1 \log h_{01}(t) + \lambda_2 \log h_{02}(t),$$

with the adjusted parameters

$$\beta_2^* = \beta_2 + \lambda_1\gamma_1 + \lambda_2\gamma_2, \qquad b_{0i}^* = \left[1 + \lambda_1\alpha_1 + \lambda_2\alpha_2\right] b_{0i}.$$

This transformation allows us to estimate β_2^* and b_{0i}^* directly, and subsequently retrieve β_2 and b_{0i}. This reparametrisation avoids identifiability issues due to the overlapping components between the longitudinal and survival processes. In particular, the overlap induced by the log-hazard terms corresponds to a linear reparametrisation of the coefficients associated with shared covariates and the random intercept.

The true values of the parameters are provided in Table 3.1 (left panel). Parameters of the graded response model are simulated by generating thresholds from a uniform distribution, which are then cumulatively summed and adjusted to ensure increasing and appropriately spaced thresholds for each item. Discrimination parameters are drawn from a log-normal distribution to ensure their positivity. This process guarantees realistic ordinal responses (comparable to HRQoL data) for the GRM framework. We set $a_1 = 1$ and $d_{1,1} = 0$ to standardise the discrimination of the first item and establish a reference point for the threshold parameters, thereby facilitating stable and interpretable estimates (Wang and Luo, 2019). For parsimony, we fix the intercept of the latent process to zero since only a single latent scale is considered and its location is already anchored by the item-level identifiability constraints. Both parametrisations, with or without a latent intercept, are commonly used in the literature. Also, as mentioned in Section 3.2.3, we consider two competing risks of dropout, denoted by $p = 1, 2$. The baseline hazard functions were specifically designed to reflect the different time

windows in which the competing dropouts (i.e. premature discontinuation of QOL completion) are most likely to occur in the context of the application of Section 3.5. Function $h_{01}(t)$ was constructed to generate a higher incidence of dropout of type 1 during the latter stages of follow-up, while $h_{02}(t)$ increases the risk of a dropout of type 2 earlier in the follow-up period. This setup was chosen to reflect the situation encountered in the application (see Section 3.5), where type 1 dropouts represent disease-unrelated dropouts, and type 2 dropouts represent disease-related dropouts. Both types of dropouts are equally represented. Each subject is censored at the end of the study, leading to a right-censoring rate of 10%. Simulated survival times were generated for every subject using the approach described in Crowther and Lambert (2013a). The cumulative hazard is evaluated using numerical quadrature, and survival times are generated using an iterative algorithm that employs the numerical root-finding method of Brent (1973).

Setting II: In this simulation setting, we generated datasets as described in Setting I, with the exception that we changed the true values of $\boldsymbol{\lambda}$. First, we set λ_1 and λ_2 to zero (Setting II.a). This allowed us to evaluate the model's ability to detect the absence of effects related to the hazards $h_{i1}(t)$ and $h_{i2}(t)$. We want to assess whether the model could correctly identify the non-significance of these effects while maintaining unbiased estimates for the other parameters. Additionally, we conducted a variant of this setting (Setting II.b), where only the effect associated with the risk of disease-unrelated dropout was non-significant ($\lambda_1 = 0$).

Setting III: In the third simulation setting, we compare our approach to a more standard classical joint model for longitudinal quality of life (QOL) data. Specifically, we examine the consequences of fitting a simplified model that does not account for the competing risk of dropout or the association between the latent trait and the logarithms of the hazard functions. The data were generated using the same model as in Setting I, but we fit the simplified model (simpleJMIRT):

$$\begin{cases} h_i(t) = h_0(t) \exp\left(\gamma\, w_i + \alpha\, b_{0i}\right) \\ \eta_i(t) = \beta_1\, t + \beta_2\, w_i + b_{0i} + b_{1i}\, t \\ \text{logit}\{\Pr(y_{ik}(t) \leq l \mid \eta_i(t))\} = d_{kl} - a_k\, \eta_i(t) \end{cases}$$

We want to assess the consequences of incorrectly specifying a model that ignores the competing risk structure of dropout and therefore the contributions of $\lambda_1 \log h_{i1}(t)$ and $\lambda_2 \log h_{i2}(t)$ in the latent process.

In each setting, to ensure convergence, we run the model estimation with

10,000 iterations, with an adaptive phase of 1,500 iterations and a burn-in of 1,500. The chains are thinned by a factor of 10, retaining 1,000 iterations for each parameter. Geweke statistics (Geweke, 1991) are used as a diagnostic tool for assessing chain convergence. For each regression parameter, we evaluate bias, root mean squared error (RMSE), and the effective coverage of 95% credible intervals (COV).

3.4.2 Simulation results

The performance measures obtained with Setting I are shown in Table 3.1 (left panel). For each regression parameter, we observe that the estimated biases are relatively low, while the coverage probabilities (COV) are close to the nominal levels. In particular, the good performances obtained for the discrimination and threshold parameters confirm that the model effectively differentiates items and their different response categories. Moreover, in the survival submodels, parameters α_1 and α_2, as well as γ_1 and γ_2, were well estimated, indicating that the model can effectively capture the distinct effects associated with the two competing risks of dropout. The same conclusion applies to parameters λ_1 and λ_2.

The results of Setting II.a can be found in Table 3.2 (left panel). We observe that the estimates of λ_1 and λ_2 are very close to zero, as expected, with minimal bias and no significant deviation from 0. Additionally, the performance in the estimation of the other longitudinal parameters β_1 and β_2 remains reliable. The results of Setting II.b are also presented in Table 3.2 (right panel) with also very satisfactory estimations of all model parameters, including λ_1 and λ_2. To further illustrate this result, Figure 3.4 (Appendix B) presents the posterior distributions of λ_1 and λ_2 for a representative simulated dataset under Setting II.a. The posterior mass is concentrated around zero, and the corresponding 95% credible intervals include the true value, supporting the ability of the model to correctly identify the absence of an effect.

The results of Setting III are presented in Table 3.1 (right panel). These were obtained using the simpleJMIRT approach, which does not account for the competing risks of dropout and estimates only a single hazard function for the dropout time, whatever its cause. The estimated value of γ is close to the average of the true values of γ_1 and γ_2, with a similar observation for the estimate of α. However, compared to Setting I, the performance of the parameter estimates in the graded response model within the longitudinal framework performs noticeably worse, exhibiting not only higher RMSE, but also lower effective coverage that falls short of the nominal 95% level, likely due to bias and model misspecification. For instance, the bias for d_{31} rises from 0.004

to 0.173, and the coverage declines from 93.0% to 27.5%. This suggests that the simplified model has difficulty providing reliable estimates, especially for threshold parameters, likely because it fails to account for the association between the latent trait and the risk of dropout.

Overall, the simulation study demonstrates the good performance of the JMIRT approach in analysing ordered categorical longitudinal data in the presence of different and possibly informative dropout types. The results confirm that the model accurately identifies the non-significance of irrelevant effects while maintaining robust performance in estimating other parameters. Furthermore, these simulations show the importance of accounting for the competing risk structure of dropouts and the potential association between the latent trait and the risks of informative dropout.

3.5 Application to glioblastoma data

The dataset used in this study originates from a phase III clinical trial involving patients with first progression of glioblastoma. This trial was conducted by the European Organisation for Research and Treatment of Cancer (EORTC-26101 trial) to determine whether the combination of lomustine and bevacizumab (combination group) improved overall survival compared to lomustine alone (monotherapy group). The details of this trial can be found at ClinicalTrials.gov (2021) (Identifier NCT01290939). While the primary endpoint of the trial is overall survival, these patients have a poor prognosis, making it crucial to evaluate the treatment's impact on their well-being. To this end, quality of life data were collected longitudinally using the EORTC Quality of Life Questionnaire (QLQ-C30) (Fayers et al., 2001) version 3, which assesses various dimensions of HRQoL. The QLQ-C30 is composed of 30 ordinal items that measure different aspects of quality of life, including a global health status/HRQoL scale, five functional scales (physical, role, emotional, social, and cognitive), three symptom scales (fatigue, nausea and vomiting, and pain), and six single items (dyspnoea, insomnia, appetite loss, constipation, diarrhoea, and financial difficulties). Patients self-reported their responses at multiple follow-up points: at baseline and every 12 weeks thereafter, allowing the monitoring of changes in HRQoL across different dimensions over time.

Trial results were already published (Wick et al., 2017) and did not demonstrate a survival advantage of treatment with lomustine plus bevacizumab over lomustine alone in patients with progressive glioblastoma, despite a prolonged progression-free survival (PFS). Additionally, the health-related quality of life was analysed using the sum-score approach, but no significant im-

Table 3.1: Simulation results for S=500 replications of sample size $n = 500$ in Setting I (left table) and Setting III (right table) with the JMIRT and the simpleJMIRT approaches respectively: bias, root mean square error (RMSE) and effective coverage (COV) of 95% credible intervals for the regression parameters. Est. gives the estimated values of survival parameters in the misspecified model.

		Setting I		
	True	Bias	RMSE	COV
a_2	0.851	-0.006	0.055	0.929
a_3	1.237	-0.008	0.082	0.930
d_{12}	1.440	-0.015	0.061	0.932
d_{13}	1.962	-0.021	0.071	0.925
d_{21}	-1.011	0.005	0.071	0.937
d_{22}	-0.466	0.003	0.070	0.925
d_{23}	0.440	-0.003	0.071	0.939
d_{31}	-1.043	0.004	0.087	0.930
d_{32}	-0.214	0.001	0.083	0.948
d_{33}	0.621	-0.009	0.089	0.942
β_1	0.150	-0.011	0.026	0.955
β_2	0.375	0.013	0.111	0.941
λ_1	-0.300	0.077	0.108	0.975
λ_2	0.200	-0.067	0.094	0.968
γ_1	-0.750	-0.016	0.139	0.932
γ_2	-1.000	-0.022	0.161	0.928
α_1	0.250	0.016	0.091	0.941
α_2	-0.250	0.001	0.100	0.941

		Setting III		
	True	Bias	RMSE	COV
a_2	0.851	-0.043	0.072	0.876
a_3	1.237	-0.060	0.105	0.843
d_{12}	1.440	0.030	0.059	0.926
d_{13}	1.962	0.027	0.070	0.924
d_{21}	-1.011	0.122	0.141	0.466
d_{22}	-0.466	0.121	0.130	0.441
d_{23}	0.440	0.116	0.132	0.457
d_{31}	-1.043	0.173	0.186	0.275
d_{32}	-0.214	0.174	0.181	0.206
d_{33}	0.621	0.168	0.176	0.262
β_1	0.150	-0.007	0.024	0.941
β_2	0.375	0.011	0.103	0.932
	Est.			
γ	-0.877	/	/	/
α	0.015	/	/	/

Table 3.2: Simulation results for S=500 replications of sample size $n = 500$ in Setting II.a (left panel) and Setting II.b (right panel) with the JMIRT approach: bias, root mean square error (RMSE) and effective coverage (COV) of 95% credible intervals for the regression parameters.

		Setting II.a		
	True	Bias	RMSE	COV
a_2	0.851	-0.006	0.065	0.936
a_3	1.237	-0.007	0.090	0.938
d_{12}	1.440	-0.013	0.060	0.919
d_{13}	1.962	-0.026	0.075	0.913
d_{21}	-1.011	0.005	0.070	0.932
d_{22}	-0.466	-0.002	0.065	0.954
d_{23}	0.440	-0.003	0.064	0.950
d_{31}	-1.043	0.004	0.081	0.952
d_{32}	-0.215	-0.001	0.079	0.943
d_{33}	0.621	-0.012	0.083	0.940
β_1	0.150	0.003	0.024	0.967
β_2	0.375	0.009	0.099	0.938
λ_1	0.000	-0.009	0.099	0.981
λ_2	0.000	0.006	0.096	0.979
γ_1	-0.750	-0.018	0.139	0.950
γ_2	-1.000	-0.020	0.161	0.930
α_1	0.250	0.012	0.091	0.940
α_2	-0.250	-0.005	0.103	0.938

		Setting II.b		
	True	Bias	RMSE	COV
a_2	0.851	-0.023	0.065	0.920
a_3	1.237	-0.028	0.092	0.922
d_{12}	1.440	-0.015	0.069	0.914
d_{13}	1.962	-0.027	0.085	0.915
d_{21}	-1.011	-0.011	0.075	0.929
d_{22}	-0.466	-0.012	0.073	0.918
d_{23}	0.440	-0.010	0.070	0.921
d_{31}	-1.043	-0.012	0.087	0.923
d_{32}	-0.214	-0.014	0.096	0.931
d_{33}	0.621	-0.017	0.098	0.941
β_1	0.150	-0.003	0.025	0.969
β_2	0.375	0.019	0.102	0.949
λ_1	0.000	0.018	0.102	0.978
λ_2	0.200	-0.027	0.097	0.976
γ_1	-0.750	-0.019	0.138	0.947
γ_2	-1.000	-0.020	0.160	0.932
α_1	0.250	0.014	0.091	0.943
α_2	-0.250	-0.001	0.101	0.949

pact of treatment was found.

To demonstrate our methodology, we focus on the physical functioning scale, which consists of the following five items:

Physical functioning scale

Item 1. *Do you have any trouble doing strenuous activities, such as carrying a heavy shopping bag or a suitcase?*

Item 2. *Do you have any trouble taking a long walk?*

Item 3. *Do you have any trouble taking a short walk outside of the house?*

Item 4. *Do you need to stay in bed or a chair during the day?*

Item 5. *Do you need help with eating, dressing, washing yourself, or using the toilet?*

Response categories:

1 = "Not at all" 2 = "A little" 3 = "Quite a bit" 4 = "Very much"

Each item has a response range from 1 ("Not at all") to 4 ("Very much"), where higher responses indicate worse physical functioning. As described in previous sections, accurately estimating the latent process requires accounting for patient dropout and distinguishing between dropouts unrelated to the disease (cause 1) and from those related to death or disease progression (cause 2). Following Cuer et al. (2022), dropout was classified as cause 2 if death or progression occurred between the last questionnaire completed and the subsequent planned visit; otherwise, it was classified as cause 1.

Because HRQoL assessments were scheduled at discrete follow-up visits every 12 weeks, we introduced a short tolerance window around the interval defined by the last completed questionnaire and the subsequent planned visit when defining disease-related dropout. This was done to account for situations in which disease progression occurred shortly before the last completed assessment or shortly after the missed visit, and could plausibly explain the subsequent non-completion of the questionnaire. Without such a tolerance window, these situations would be classified as disease-unrelated dropout, since no event occurs strictly between the last completed questionnaire and the dropout time. However, this would ignore the delayed or slightly shifted impact of clinical deterioration due to the discrete timing of HRQoL measurements. Therefore, in such cases, dropout was classified as cause 2 rather than cause 1, as this better reflects the underlying clinical mechanism.

The dataset analysed includes $n = 423$ patients enrolled across multiple institutions in Europe. Of these, 145 were randomised to receive monotherapy

treatment, while the rest were randomised to receive a combination of lomustine and bevacizumab. Follow-up times for the longitudinal HRQoL data ranged from 1 to 96 weeks. The median overall survival was 40.4 weeks, and the median progression-free survival was 13.6 weeks (see Appendix B, Figure 3.2). Disease-unrelated dropout (cause 1) occurred in 31.2% of patients, while 32.4% experienced disease-related dropout (cause 2). The remaining patients, who did not drop out, were right-censored at the end of the study period. The cumulative incidence functions are shown in Figure 3.3 (Appendix B). Each patient completed the corresponding HRQoL questionnaire between 1 and 9 times during their follow-up.

We consider the following joint model in our analysis:

$$\begin{cases} h_i^{\text{unr}}(t) = h_0^{\text{unr}}(t) \exp\left(\gamma_{\text{Trt}}^{\text{unr}}\, w_{\text{Trt}} + \alpha^{\text{unr}}\, b_{0i}\right) \\ h_i^{\text{rel}}(t) = h_0^{\text{rel}}(t) \exp\left(\gamma_{\text{Trt}}^{\text{rel}}\, w_{\text{Trt}} + \alpha^{\text{rel}}\, b_{0i}\right) \\ \eta_i(t) = \boldsymbol{\beta}_1\, t + \boldsymbol{\beta}_2\, w_{\text{Trt}} + b_{0i} + b_{1i}\, t + \boldsymbol{\lambda}_1 \log h_i^{\text{unr}}(t) + \boldsymbol{\lambda}_2 \log h_i^{\text{rel}}(t) \\ \text{logit}\{\Pr(y_{ik}(t) \le l \mid \eta_i(t))\} = d_{kl} - a_k\, \eta_i(t) \end{cases}$$

where η_{ij} represents the latent physical functioning of patient i at time j, with $l = 1, \ldots, 4$ indicating the response categories and $k = 1, \ldots, 5$ representing the five items of the physical functioning scale. Larger values of η_{ij} correspond to worse physical functioning, while lower values indicate better functioning. The covariate w_{Trt} denotes the treatment indicator, taking the value 0 for monotherapy and 1 for combination therapy. The superscripts "unr" and "rel" refer to disease-unrelated dropout (cause 1) and disease-related dropout (cause 2), respectively.

As presented in Section 3.3, we used Bayesian methods with Markov Chain Monte Carlo techniques to estimate this joint model. Trace plots revealed no discernible trends in the MCMC chains, suggesting adequate convergence. Furthermore, the Gelman-Rubin diagnostic confirmed that the potential scale reduction factor (R) for all parameters remained below 1.1, further validating convergence. The JMIRT approach takes 6.31 minutes on a single core of a 2.6 GHz Intel Core i5-1145G7 processor, benefiting from a combination of R and C++ code for enhanced speed. The estimated parameters are presented in Table 3.3, with the left panel reporting the survival and latent process parameters, and the right panel reporting the GRM parameters.

Coefficients β_1 and β_2 represent the effects of time and treatment, respectively, on the latent trait $\eta_i(t)$, which reflects physical functioning. The significant positive estimate of β_1 suggests that, over the course of up to 96 weeks of

Table 3.3: Estimation results of survival submodels and latent process parameters (left panel) and GRM submodel parameters (right panel) with the JMIRT approach: estimates and 95% credible intervals (CI) are presented.

		Est.	CI 95%
β_1	(t_{ij})	0.050	[0.035; 0.065]
β_2	(w_{trt})	0.447	[-0.208; 1.018]
λ_1	$(\log h_i^{\text{unr}}(t_{ij}))$	-0.290	[-0.478;-0.051]
λ_2	$(\log h_i^{\text{rel}}(t_{ij}))$	0.120	[0.027; 0.245]
$\gamma_{\text{Trt}}^{\text{unr}}$	(w_{trt})	-0.346	[-0.807; 0.122]
$\gamma_{\text{Trt}}^{\text{rel}}$	(w_{trt})	-0.717	[-1.111;-0.351]
α^{unr}	(b_{0i})	0.192	[0.097; 0.284]
α^{rel}	(b_{0i})	0.024	[-0.086; 0.122]

		Est.	CI 95%
	a_1	1.000	/
	a_2	1.127	[0.944; 1.503]
	a_3	0.993	[0.852; 1.235]
	a_4	0.630	[0.544; 0.743]
	a_5	0.969	[0.816; 1.195]
Item1	d_{11}	0.000	/
	d_{12}	3.384	[2.740; 3.911]
	d_{13}	6.027	[4.952; 6.858]
Item2	d_{21}	0.301	[0.018; 0.594]
	d_{22}	2.645	[2.254; 3.012]
	d_{23}	5.100	[4.538; 5.624]
Item3	d_{31}	2.615	[2.267; 2.991]
	d_{32}	5.020	[4.564; 5.415]
	d_{33}	6.781	[6.297; 6.991]
Item4	d_{41}	0.591	[0.396; 0.782]
	d_{42}	2.770	[2.472; 3.034]
	d_{43}	4.813	[4.313; 5.314]
Item5	d_{51}	3.778	[3.382; 4.135]
	d_{52}	5.348	[4.848; 5.843]
	d_{53}	6.647	[6.080; 6.979]

follow-up, physical functioning tends to deteriorate. Additionally, the combined treatment does not appear to have a significant effect on $\eta_i(t)$, suggesting that it may not impact the physical ability of patients. A similar result was observed in the original presentation of the study results (Wick et al., 2017). Parameters λ_1 and λ_2 capture the influence of the log-transformed hazard functions for each type of dropout on the latent trait over time. The significance of both estimates underscores the importance of incorporating risk-related information in the specification of the latent trait, a process made feasible by our approach JMIRT. Specifically, $\hat{\lambda}_2 = 0.120$ indicates that, at a given time point, a higher risk of disease-related dropout is associated with worse physical functioning at the same time point, while a higher risk of disease-unrelated dropout is associated with better physical functioning ($\hat{\lambda}_1 = -0.290$). This latter finding may suggest that patients who experience improvements in their well-being may choose to discontinue follow-up visits because they no longer perceive a need for continued monitoring.

The estimated discrimination parameters ($\hat{a}_k$) range from 0.630 for Item 4 to 1.127 for Item 2 (see Table 3.3). Overall, Items 1, 2, 3, and 5 display comparable levels of discrimination, whereas Item 4 stands out with a clearly lower value. Its credible interval is the only one that does not overlap with those of the other items. Posterior comparisons between discrimination parameters (see Table 3.4 in Appendix B) confirm this pattern: all pairwise differences involving Item 4 indicate that it is consistently less discriminative than the other items. In contrast, differences among Items 1, 2, 3, and 5 remain limited. Although a_2 tends to exceed a_3 and a_5, the magnitude of these differences remains modest. This pattern is consistent with the content of the items. Item 4, which captures the need to stay in bed or a chair during the day, reflects a low level of physical functioning and provides limited discrimination across patients. In contrast, the other items assess activities requiring higher levels of physical effort, resulting in a greater ability to differentiate between patients with different levels of physical functioning.

Concerning the threshold parameters, d_{kl} represents the cut-off on the latent trait $\eta_i(t)$ at which the probability of responding in category l or below equals 50%, for item k. For example, for Item 2, a patient with a latent trait value of $\eta_i = 2.347$ (i.e., $\hat{d}_{22}/\hat{a}_2$) has a 50% probability of endorsing a response of 2 ("A Little") or lower. Higher d_{kl} values indicate that more severe functioning problems (higher $\eta_i(t)$ values) are needed to endorse higher categories of response.

As an illustration, Figure 3.1 presents the estimated distribution of response probabilities for Item 2 (*Do you have any trouble taking a long walk?*). Specif-

ically, it illustrates the temporal evolution of the probabilities that an average patient ($\mathbf{b}_i = (0, 0)$) selects a specific response category $l \in \{1, \ldots, 4\}$ at various time points $t \in \{10, 20, 30, 40, 50, 60\}$ (in weeks) after randomization, while accounting for the two types of dropout. As the treatment does not have a significant effect on η_{ij}, this covariate was taken out of the model. The black dots represent the estimated values of $\eta_i(t)$ for such a patient. At time $t = 10$, the patient is most likely to select the response category "A Little" for Item 2, with $\mathrm{P}(Y_{10} = 2) = 52.6\%$. Over time, as the value of η increases, the probability of selecting higher response categories rises, with "Quite a Bit" becoming the most likely category starting at $t = 40$.

Regarding the survival submodel parameters, the treatment demonstrates a significant negative effect ($\hat{\gamma}^{\text{rel}}_{\text{Trt}} = -0.717$) on the risk of disease-related dropout, suggesting that patients receiving the combination treatment are less likely to withdraw due to death or progression of the disease compared to those receiving lomustine alone. This is consistent with previous findings (Wick et al., 2017) showing that the combination treatment prolonged progression-free survival (PFS). In contrast, treatment has no significant effect on the risk of disease-unrelated dropout (see $\hat{\gamma}^{\text{unr}}_{\text{Trt}}$ in Table 3.3). This distinction shows the importance of understanding the reasons behind dropout. Moreover, the association between the individual-specific deviation from the latent trait and the risks of dropout is significant only in the disease-unrelated case ($\hat{\alpha}^{\text{unr}} = 0.192$).

We also note the importance of distinguishing between the interpretations of α^{unr} and λ_1 (and similarly for α^{rel} and λ_2). While both parameters assess the association between the risk of dropout and the latent trait, they approach it from different perspectives: one focuses on individual-specific effects, while the other examines the impact of the dropout hazard on the latent trait. Together, they account for the complexity of the dropout process in longitudinal studies, especially in the context of assessing quality of life.

3.6 Discussion

The primary aim of this study was to develop the JMIRT approach to efficiently analyse multiple longitudinal ordinal categorical data while accounting for potential informative dropout. This methodology is framed within a joint modelling approach that connects a latent variable, derived from multiple longitudinal ordinal outcomes in an IRT model, to cause-specific hazards of dropout. The latent variable is regressed on random effects and predictors, including survival information. Unlike traditional joint models, which treat longitudinal data as a covariate in the survival submodel, our approach priori-

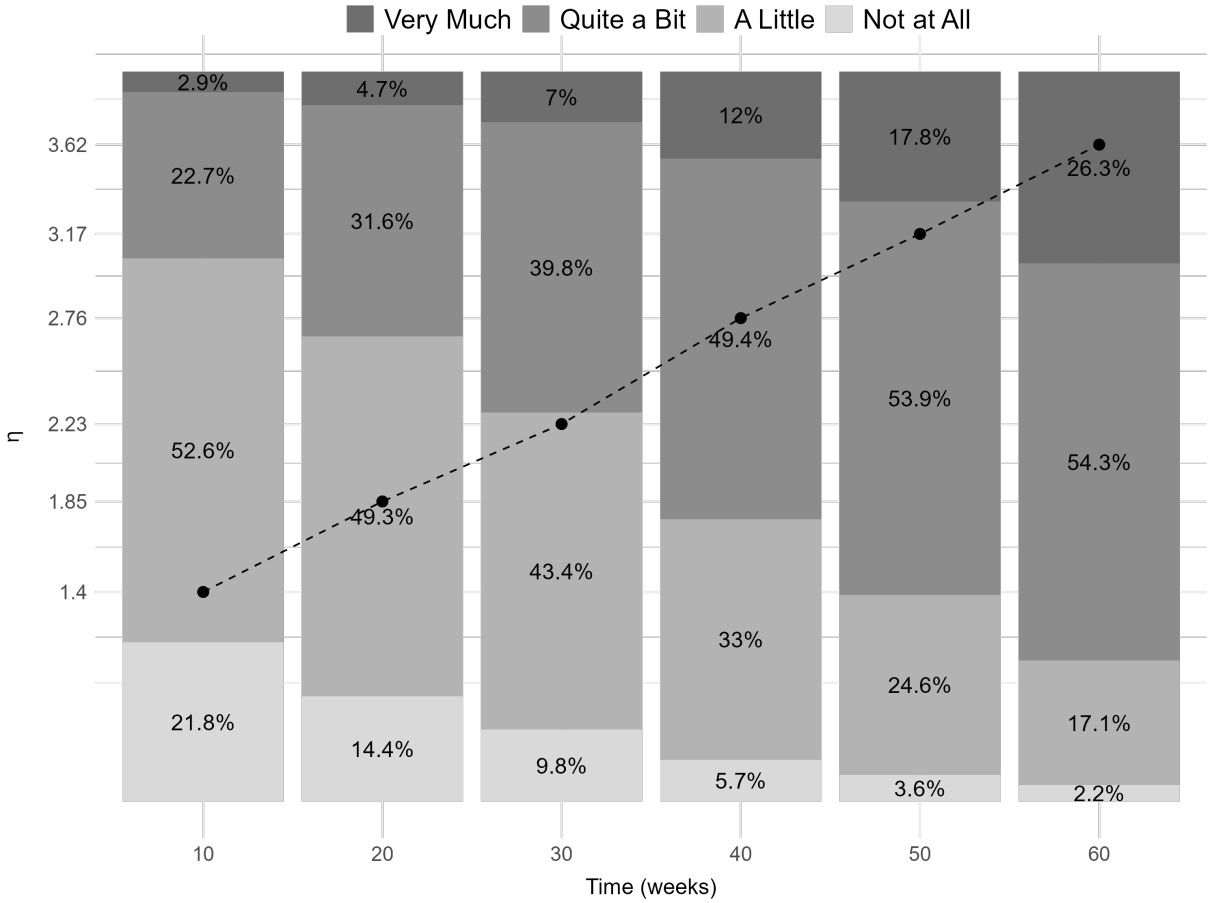

Figure 3.1: Distribution of response probabilities over time for *Item 2*. The stacked bars show the probabilities of selecting each response category at different times. Black dots represent the estimated values of $\eta_i(t)$ for an average patient i.

tizes the longitudinal components by incorporating the log risks of both competing dropout events as covariates in the latent process model. The JMIRT approach thus provides a novel and robust tool for analysing multiple ordinal categorical data, particularly in health-related quality of life assessments in the presence of (potentially informative) dropout. We note that, as in any joint modelling framework, inference relies on the adequate specification of both the longitudinal and survival submodels.

The simulation studies showed that the proposed inference strategy effectively estimates the model parameters in an JMIRT model with minimal bias. In the first setting, we demonstrated the model's ability to differentiate between response categories and accurately estimate the distinct effects associated with the two competing risks of dropout. In Setting II, the model successfully detected the absence of dropout effects on longitudinal ordinal outcomes, while maintaining accurate estimation of other parameters. Setting III highlights the importance of accounting for patient dropout and distinguishing between its different causes to obtain accurate estimates of the latent process underlying the multivariate longitudinal ordinal responses.

In our application to glioblastoma data, the JMIRT approach provided a comprehensive analysis of HRQoL data and was able to capture the impact of disease-related dropout on patients' physical functioning over time. Our findings revealed that a higher risk of disease-related dropout was associated with worse physical functioning, whereas disease-unrelated dropout was linked to better functioning. This underscores the necessity of accounting for different dropout mechanisms when analysing longitudinal quality of life data in clinical trials.

While these results highlight the relevance of the proposed approach, limitations related to the classification of dropout should be noted.

Studying the timing and reasons for patient dropout is critical, as it provides valuable information that can influence the analysis and interpretation of longitudinal data. However, accurately classifying the cause of dropout remains challenging and cannot be considered fully certain in practice. Although some clinical trials have started to collect information on the reasons for questionnaire non-completion, this information is often incomplete or inconsistently recorded, particularly in older datasets. As a result, we could not rely exclusively on the recorded dropout categories and instead defined a study-specific classification rule, as described in Section 3.5. We note that this classification was broadly consistent with the reported reasons for questionnaire non-completion when such information was available. Nevertheless, it remains

a constructed approximation and may not fully reflect the underlying clinical reality. This limitation highlights the inherent difficulty of distinguishing between disease-related and unrelated dropout in real-world data and may introduce some degree of misclassification.

Despite these limitations, the proposed approach provides a structured framework to account for dropout mechanisms and contributes to advancing research on this topic. As demonstrated in this chapter, ignoring dropout or failing to distinguish between its different causes may lead to biased estimates and an incomplete understanding of the true effects of covariates. Methodological frameworks such as the JMIRT approach, which explicitly account for both dropout timing and dropout cause, underline the importance of these distinctions. We hope that such developments will encourage more systematic and precise recording of dropout circumstances in future clinical trials, thereby improving the reliability of longitudinal HRQoL analyses.

From a modelling perspective, several assumptions underlying the IRT component also deserve discussion. In practice, the adequacy of IRT assumptions can be evaluated using standard psychometric diagnostics. These include the assessment of dimensionality (e.g., through exploratory or confirmatory factor analysis (De Ayala, 2013)), the investigation of local item dependence (e.g., residual-based diagnostics), and the evaluation of measurement invariance across groups or over time (e.g., differential item functioning (Holland and Wainer, 2012)). In the present work, the primary objective is methodological, and a full psychometric validation of the questionnaire was beyond the scope of the analysis. However, the use of the well-established QLQ-C30 questionnaire supports the plausibility of the modelling assumptions. Nevertheless, potential violations of these assumptions, such as lack of measurement invariance over time or response shift, cannot be excluded and may challenge the stability of item parameters. Model fit can also be examined through posterior predictive checks in a Bayesian framework (Gelman et al., 2013), which allow for assessing how well the model generates data that resemble the observed questionnaire responses. In addition, information criteria such as the Watanabe–Akaike information criterion (WAIC) provide a robust framework for comparing the performance of the proposed joint IRT model against competing specifications or simpler alternatives, such as linear mixed models applied to longitudinal sum-scores (see Section 1.2.3).

Within the framework of joint models, several extensions can be considered. Although we specifically accounted for competing dropout events, this methodology can accommodate other competing types of events of interest. In this chapter, we modelled the categorical ordinal data using a graded re-

sponse model, but other models within the IRT framework could easily be employed, depending on the structure of the raw data (Van der Linden, 2016). Also, while our application focused on HRQoL data, the JMIRT approach is not limited to this context and can be applied to any longitudinal ordinal categorical data, including other types of patient-reported outcome measures (PROMs). Additionally, more complex, non-linear trajectories in the evolution of the latent trait could be captured using spline-based approaches. Finally, we assumed that the multiple items studied reflect a univariate latent variable. Exploring a multidimensional latent trait model can provide valuable insights into different aspects of multivariate ordinal outcomes (Wang and Luo, 2019). However, adding multiple longitudinal latent traits to capture these dimensions would substantially increase the computational complexity.

3.7 Appendix B

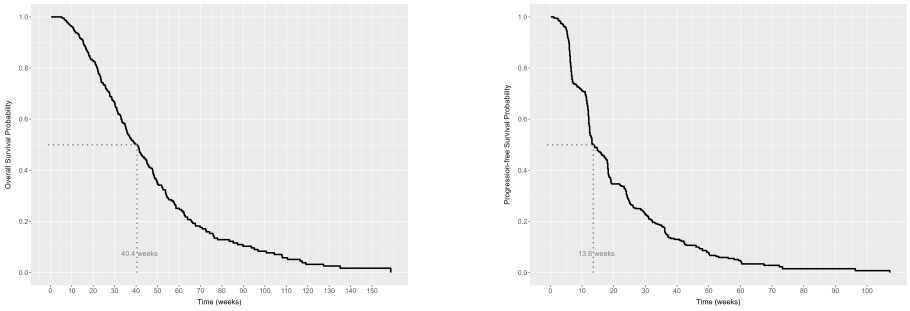

Figure 3.2: Overall Survival (left panel) and Progression-free Survival (right panel) in patients with first progression of glioblastoma

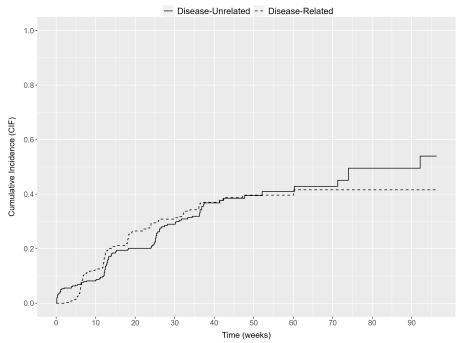

Figure 3.3: Cumulative incidence functions for dropouts unrelated and related to disease progression or death.

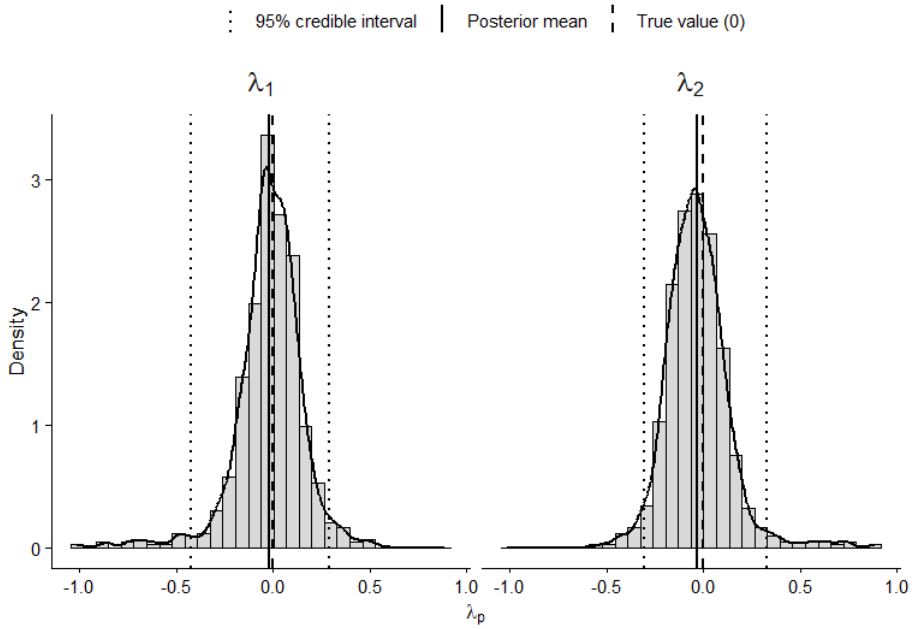

Figure 3.4: Posterior distribution of λ_1 and λ_2 under Setting II.a for a representative simulated dataset. The histogram and overlaid kernel density estimate are based on the MCMC posterior sample.

Table 3.4: Posterior pairwise comparisons of discrimination parameters. Median posterior differences, 95% credible intervals (CI), and posterior probabilities are reported. A positive difference indicates a higher discrimination for the first item in the comparison.

Comparison	Median	CI 95%	P(> 0)
$a_1 - a_2$	-0.09	[-0.50 ; 0.06]	0.14
$a_1 - a_3$	0.02	[-0.23 ; 0.15]	0.57
$a_1 - a_4$	0.37	[0.26 ; 0.45]	1.00
$a_1 - a_5$	0.04	[-0.20 ; 0.18]	0.69
$a_2 - a_3$	0.13	[0.01 ; 0.31]	0.99
$a_2 - a_4$	0.48	[0.35 ; 0.77]	1.00
$a_2 - a_5$	0.15	[0.01 ; 0.35]	0.99
$a_3 - a_4$	0.35	[0.25 ; 0.52]	1.00
$a_3 - a_5$	0.03	[-0.08 ; 0.14]	0.64
$a_5 - a_4$	0.33	[0.22 ; 0.49]	1.00

A slope-corrected two-stage joint model for multidimensional ordinal HRQoL and survival data

4

Health-related quality-of-life (HRQoL) outcomes are increasingly incorporated into oncology research to complement traditional survival endpoints by capturing patients' well-being over time. These outcomes are typically collected through multidimensional questionnaires yielding longitudinal ordinal data, and are often subject to dropout due to disease progression or death. In this context, joint models provide a well-established framework to account for the dependence between longitudinal HRQoL trajectories and time-to-event outcomes, but fully joint estimation rapidly becomes computationally prohibitive when multiple latent dimensions and random effects are involved. In this chapter, we propose a novel slope-corrected two-stage (SC2S) approach for the joint analysis of multivariate ordinal HRQoL data and survival outcomes within a multidimensional latent trait framework. The proposed approach propagates longitudinal information to the survival model through informative priors on the random effects, while additionally re-estimating longitudinal slope parameters. This strategy substantially reduces bias in both longitudinal and survival submodels while preserving much of the computational efficiency of two-stage procedures. Through simulation studies and an application to HRQoL data from patients with progressive glioblastoma, we show that the proposed method closely approximates fully joint Bayesian estimation while requiring notably less computation time.

The material presented in this chapter is based on the paper: Doms H, Lambert P, Legrand C. Bias-Corrected Two-Stage Modeling for Joint Analysis of Multidimensional HRQoL and Survival, submitted to *Pharmaceutical Statistics*.

4.1 Introduction

In oncology clinical research, overall survival (OS) is traditionally regarded as the primary endpoint for assessing the benefit of new treatments. Although OS remains a key measure of therapeutic efficacy, it does not fully reflect how treatments affect patients' daily functioning and well-being. Consequently, complementary patient-centered endpoints have emerged, particularly health-related quality of life (HRQoL). HRQoL is typically assessed using patient-reported questionnaires administered repeatedly throughout treatment and follow-up, offering valuable insight into patients' perceptions of their physical, psychological, and social health (Fayers and Machin, 2013). Patient-reported outcomes such as HRQoL thus provide essential complementary information, contributing to a more comprehensive and patient-centred assessment of treatment benefit. Their importance has also motivated the development of advanced statistical methods to appropriately analyse these complex longitudinal data.

HRQoL questionnaires are typically composed of multiple scales, each assessed through several items. Responses are often recorded on Likert-type scales (e.g., "not at all," "a little," "quite a bit," "very much"), producing ordinal data. A widely used method for analyzing such data is the scoring approach (Fayers et al., 2001), in which item responses are summed to obtain a total score for each scale and time point. However, this method presents important limitations (Gorter et al., 2015): it ignores heterogeneity in response patterns that yield the same total score, leading to loss of information, and the resulting scores are frequently treated as continuous despite their ordinal nature and potential ceiling or floor effects.

To overcome these issues, item response theory (IRT) methods, and in particular graded response (GRM) models, have been increasingly adopted for HRQoL analysis (Barbieri et al., 2017). These models estimate a latent variable that represents the unobserved true health status using observed item responses regarded as manifestations of this latent construct. To capture the correlation structure inherent to longitudinal data, multilevel IRT extensions have been proposed, allowing the latent trajectory to depend on covariates and subject-specific random effects. More recently, the unidimensional assumption has been relaxed by introducing multidimensional latent trait mixed models (Wang and Luo, 2017), enabling multiple latent dimensions and, when relevant, within-item multidimensionality.

These longitudinal data are frequently analysed in the presence of intercurrent or terminal events. In clinical trials, patients may experience events such

as death or loss to follow-up, which interrupt the collection of HRQoL measurements. The occurrence of such events is often related to the underlying health status, resulting in dependence between the event process and the longitudinal outcomes. Ignoring this informative dropout may therefore lead to biased estimates (Henderson et al., 2000). In this context, the joint analysis of the event hazard and longitudinal data, known as the joint modelling framework, has become increasingly common. Touraine et al. (2023) recently highlighted the importance of joint models for HRQoL data when informative dropout is present. Saulnier et al. (2022) extended the joint modelling methodology to handle repeated data from measurement scales using a cumulative probit model and cause-specific hazard functions. Doms et al. (2025) proposed an approach to handle informative dropout in longitudinal HRQoL data through a joint modelling framework, which integrates a graded response model for multivariate ordinal outcomes and a cause-specific proportional hazards model for informative and non-informative dropout events. Both of these approaches are restricted to a unidimensional latent trait. However, Wang and Luo (2019) introduced a joint multidimensional latent trait mixed model with a proportional hazards submodel in a Bayesian framework. Because they rely on fully joint estimation, these methods are computationally intensive, particularly when several latent dimensions and random effects are involved. Indeed, joint models for multivariate longitudinal data often face substantial computational burden, as the number of random effects grows with the number of outcomes, leading to slow estimation algorithms (Papageorgiou et al., 2019).

Early approaches to the estimation of joint models relied on two-stage methods (Self and Pawitan, 1992; Tsiatis et al., 1995). In these approaches, the longitudinal submodel is fitted first, and its estimates are subsequently used in the survival model. Although computationally appealing, such procedures may produce biased estimates under informative dropout (Tsiatis and Davidian, 2004; Ye et al., 2008). For this reason, full joint estimation was later proposed. This approach corrects bias by estimating both submodels simultaneously, but it is substantially more computationally demanding. As joint models have evolved to accommodate increasingly complex longitudinal and survival structures, their computational burden has grown accordingly. To address these challenges, some authors have proposed bias-corrected two-stage methods, which retain the computational advantages of two-stage estimation while reducing bias. Mauff et al. (2020) introduced importance-sampling corrections to reduce bias in multivariate joint models, while Alvares and Leiva-Yamaguchi (2023) extended the Bayesian two-stage framework with informative priors for random effects and bias-adjustment techniques, improving accuracy while maintaining computational efficiency. Both of these approaches

are developed within the context of exponential-family distributions for the longitudinal responses.

In this work, we adapt the corrected two-stage approach proposed by Alvares and Leiva-Yamaguchi (2023) to efficiently estimate a joint model combining a multidimensional latent trait mixed model for multiple longitudinal ordinal outcomes and a proportional hazards model with a flexible baseline hazard, within a Bayesian framework. In addition, we propose a correction for the longitudinal slope parameters, mitigating the bias that may arise when the longitudinal and survival components are fitted separately. Our goal is to approximate the accuracy of fully joint estimation while significantly reducing the computational cost.

The rest of this chapter is organized as follows. Section 4.2 introduces the proposed joint model. Section 4.3 presents the Bayesian estimation strategy based on the corrected two-stage approach. Section 4.4 reports a simulation study assessing the method's performance, and Section 4.5 applies our approach to HRQoL data from the QLQ-BN20 questionnaire completed by glioblastoma patients. Section 4.6 concludes the chapter.

4.2 Methodology

4.2.1 Longitudinal submodel

We consider a sample of n patients followed over time. For each individual $i = 1, \ldots, n$, we observe K longitudinal ordinal items, where item k ($k = 1, \ldots, K$) has L_k ordered categories. The longitudinal measurements are denoted by $y_{ik}(t)$, corresponding to item k observed at time $t = t_{ij}$ ($j = 1, \ldots, N_i$). All items are coded such that larger values indicate worse clinical conditions.

We assume the existence of P latent variables, also called latent dimensions, with $P < K$, representing distinct abilities for each subject. The vector of latent processes for individual i at time t is denoted as

$$\boldsymbol{\eta}_i(t) = \left(\eta_i^{(1)}(t), \ldots, \eta_i^{(p)}(t), \ldots, \eta_i^{(P)}(t)\right)^\top.$$

We consider that each item loads on one or more latent dimensions and that higher values of $\eta_i^{(p)}(t)$ correspond to a lower ability level for dimension p.

4.2.1.1 Latent process

Latent variables cannot be directly observed but can be assessed through the longitudinal responses provided to the different items. The trajectory of $\boldsymbol{\eta}$

over time for patient i is described using a linear mixed model. Specifically, for the pth latent variable ($p = 1, \ldots, P$), we have :

$$\begin{cases} \eta_i^{(p)}(t) = \mathbf{x}_i^{(p)}(t)^\top \boldsymbol{\beta}^{(p)} + \mathbf{z}_i^{(p)}(t)^\top \mathbf{b}_i^{(p)} \\ \mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}) \end{cases}$$

where $\mathbf{x}_i^{(p)}(t)$ and $\mathbf{z}_i^{(p)}(t)$ are the covariates associated with the fixed and random effects for each latent variable $\eta_i^{(p)}(t)$. The vector $\mathbf{b}_i = (\mathbf{b}_i^{(1)}, \ldots, \mathbf{b}_i^{(P)})^\top$ contains all the normally distributed individual random effects. The correlations between latent variables are captured through the covariance structure of $\mathbf{b}_i$. The fixed effects term describes the average trajectory of the latent process at the population level, while the random effects term captures individual-specific deviations from this trajectory.

4.2.1.2 Graded response model

We now define the link between the latent process $\boldsymbol{\eta}_i(t)$ and each observed longitudinal ordinal response using the graded response model (GRM) (Samejima, 1969).

The cumulative probability for the kth ordinal outcome, with L_k ordered categories, is defined for $l = 1, \ldots, L_k - 1$ as:

$$\begin{aligned} &\text{logit}\{\Pr(y_{ik}(t) \leq l \mid \boldsymbol{\eta}_i(t))\} = d_{kl} - \boldsymbol{a}_k^\top \boldsymbol{\eta}_i(t) \\ \Leftrightarrow \quad &\Pr(y_{ik}(t) \leq l \mid \boldsymbol{\eta}_i(t)) = \frac{1}{1 + \exp\{-(d_{kl} - \boldsymbol{a}_k^\top \boldsymbol{\eta}_i(t))\}} \end{aligned}$$

Each item k is characterized by a vector of discrimination parameters $\boldsymbol{a}_k = (a_k^{(1)}, \ldots, a_k^{(P)})^\top$ of dimension P, and by a set of threshold parameters $d_k = (d_{k1}, \ldots, d_{k,L_k-1})^\top$. The discrimination vector $\boldsymbol{a}_k$ quantifies the strength of association between item k and each latent dimension. A larger absolute value of $a_k^{(p)}$ indicates that item k is more sensitive to differences along latent dimension p. The threshold parameter d_{kl} determines, together with $\boldsymbol{a}_k$, the location in the latent space where the probability of responding in category l or below equals 50%. Specifically, this occurs for latent variable vectors $\boldsymbol{\eta}_i(t)$ satisfying $\boldsymbol{a}_k^\top \boldsymbol{\eta}_i(t) = d_{kl}$. The thresholds are subject to the ordering constraint

$$d_{k1} < d_{k2} < \cdots < d_{k,L_k-1}.$$

The probability of observing a response in category $l = 1, \ldots, L_k$ is obtained from the cumulative probabilities as

$$\Pr(y_{ik}(t) = l \mid \boldsymbol{\eta}_i(t)) = \Pr(y_{ik}(t) \leq l \mid \boldsymbol{\eta}_i(t)) - \Pr(y_{ik}(t) \leq l - 1 \mid \boldsymbol{\eta}_i(t)),$$

with the standard conventions

$$\Pr(y_{ik}(t) \leq 0 \mid \boldsymbol{\eta}_i(t)) = 0, \qquad \Pr(y_{ik}(t) \leq L_k \mid \boldsymbol{\eta}_i(t)) = 1.$$

In multidimensional graded response models, identifiability constraints are required because the latent traits are not directly observed. Without such constraints, the model is invariant to translations, rescalings, and rotations of the latent space, meaning that different combinations of thresholds, discrimination parameters, and latent variables can produce exactly the same response probabilities. These constraints ensure identifiability by fixing, for each dimension, a reference origin and unit, and by preventing the different dimensions from overlapping. To achieve this, several identification strategies are possible, but a common approach consists in selecting, for each latent dimension, an anchor item whose parameters are constrained to fix the scale, location, and orientation of the latent space. This typically involves fixing one discrimination parameter per dimension (e.g., $a_k^{(p)} = 1$), setting one threshold to a reference value (e.g., $d_{k1} = 0$), and restricting cross-loadings for these anchor items.

4.2.2 Survival submodel

Let T_i^* denote the time to a terminal event for the i-th individual (e.g., death or disease progression). We observe $T_i = \min(T_i^*, C_i)$, where C_i is the administrative (and non-informative) right censoring time (end of study).

We model the time-to-event using a proportional hazard model

$$h_i(t) = h_0(t) \exp\left\{\boldsymbol{\gamma}^\top \mathbf{w}_i + \boldsymbol{\alpha}^\top m(t, \mathbf{b}_i)\right\}, \qquad t > 0$$

where $h_0(t)$ denotes the baseline hazard and $\mathbf{w}_i$ is the vector of baseline survival covariates for subject i with regression coefficients $\boldsymbol{\gamma}$. Parameters $\boldsymbol{\alpha}$ quantify the association between the longitudinal latent process and the survival process. Different association functions $m(t, \mathbf{b}_i)$ can be specified. For example, the current-value $m(t, \mathbf{b}_i) = \boldsymbol{\eta}_i(t)$ and the random-effects $m(t, \mathbf{b}_i) = \mathbf{b}_i$ structures are the most frequently used in the literature (Rizopoulos, 2012a), although other specifications are also possible.

In this joint modelling framework, the baseline hazard is parameterized using penalized B-splines. Therefore we define the logarithm of the baseline hazard functions using B-splines (Lambert and Eilers, 2005; Rizopoulos, 2016) :

$$\log\{h_0(t)\} = \sum_{u=1}^{U} \gamma_{h_0,u} B_u(t), \qquad \boldsymbol{\gamma}_{h_0} = (\gamma_{h_0,1}, \ldots, \gamma_{h_0,U})^\top,$$

where $\gamma_{h_0,u}$ is the regression coefficient associated to the uth function of a large B-spline basis associated to equidistant knots on $(0, \max_i T_i)$.

4.3 Bayesian inference

We apply a Bayesian approach to estimate our proposed joint model that links a time-to-event outcome and a multidimensional latent trait linear mixed model, defined by

$$\begin{cases} h_i(t) = h_0(t) \exp\left\{\boldsymbol{\gamma}^\top \mathbf{w}_i + \boldsymbol{\alpha}^\top m(t, \mathbf{b}_i)\right\}, \\ \eta_i^{(p)}(t) = \mathbf{x}_i^{(p)}(t)^\top \boldsymbol{\beta}^{(p)} + \mathbf{z}_i^{(p)}(t)^\top \mathbf{b}_i^{(p)} \\ \text{logit}\{\Pr(y_{ik}(t) \leq l \mid \boldsymbol{\eta}_i(t))\} = d_{kl} - \boldsymbol{\alpha}_k^\top \boldsymbol{\eta}_i(t) \end{cases} \tag{4.1}$$

4.3.1 Posterior distribution

As introduced in Section 4.1, the model in (4.1) may be estimated using a full joint modelling approach or via two-stage strategies. Each approach leads to a different formulation of the posterior distribution, which we detail in the following subsections. For each approach, the posterior distribution is explored using Markov Chain Monte Carlo (MCMC) methods.

4.3.1.1 Joint specification (JS)

In the framework of joint models for longitudinal and survival data, the expression for the posterior distribution of the model parameters is derived under the assumption that the longitudinal and survival processes are independent conditionally on the random effects $\mathbf{b}_i$. Moreover, the longitudinal measurements $\mathbf{y}_i$ of each individual are assumed independent given the random effects. Therefore, the joint posterior distribution is defined by

$$\begin{aligned} p(\theta, \mathbf{b} \mid \mathbf{T}, \delta, \mathbf{y}) \;\propto\; & \prod_{i=1}^{n} p(T_i, \delta_i \mid \mathbf{b}_i; \theta_s) \\ & \times \prod_{j=1}^{N_i} \prod_{k=1}^{K} p\big(y_{ik}(t_{ij}) \mid \mathbf{b}_i; \theta_y\big)\, p(\mathbf{b}_i; \theta_b)\, p(\theta). \end{aligned} \tag{4.2}$$

where θ denotes the full parameter vector with joint prior $p(\theta)$. Specifically, $\theta = (\theta_s, \theta_y, \theta_b)$, with θ_s for the survival submodel, θ_y for the longitudinal submodel, and θ_b for the random-effects covariance structure. The expressions for $p(T_i, \delta_i \mid \mathbf{b}_i; \theta_s)$, $p(y_{ik}(t) \mid \mathbf{b}_i; \theta_y)$ and $p(\mathbf{b}_i; \theta_b)$ can be found in Appendix C1.

4.3.1.2 Standard Two-Stage (S2S) and Corrected Two-Stage (C2S) approaches

EExploration of the joint posterior distribution in Equation (4.2) using MCMC is computationally intensive and may suffer from convergence issues due to the complex hierarchical structure of joint models. Some authors have proposed avoiding these limitations by using two-stage estimation strategies, as introduced in Section 4.1.

In the Standard Two-Stage (S2S) approach, the longitudinal and survival submodels are estimated sequentially. In the first stage, the longitudinal submodel is fitted independently, providing posterior estimates (often posterior means) for the parameters θ_y and individual random effects $\mathbf{b}_i$, denoted $\hat{\theta}_y$ and $\hat{\mathbf{b}}_i$. In the second stage, these estimated quantities are inserted as fixed covariates into the survival submodel, through $m_i(t|\hat{\mathbf{b}}_i, \hat{\theta}_y)$, to infer on θ_s. While computationally attractive, this approach ignores the informative dropout mechanism in the first step, leading to biased estimates. Using these biased predictors in the survival model further propagates this bias, especially for the association parameter. Moreover, by treating random effects ($\hat{\mathbf{b}}_i$) as fixed values, it implicitly assumes they are known without error in the second stage. Such an omission fails to propagate the sampling variability from the first stage, causing a significant loss in the variability of the random effects and leading to an artificial reduction in the uncertainty of the survival estimators (Tsiatis and Davidian, 2004; Ye et al., 2008).

To address these issues, Alvares and Leiva-Yamaguchi (2023) proposed a Corrected Two-Stage (C2S) approach, which provides a practical compromise between the S2S and JS approaches. As in the first stage of the S2S approach, the longitudinal submodel is fitted separately to obtain $\hat{\theta}_y$ and $\hat{\theta}_b$. In the second stage, instead of fixing $\mathbf{b}_i$ to their point estimates after the first step, the posterior distribution of $(\theta_s, \mathbf{b})$ is derived using the likelihood of the joint model, with the longitudinal parameters and the variance–covariance matrix of the random effects treated as known and replaced by their estimates from the first step:

$$\begin{aligned} p(\theta_s, \mathbf{b} \mid \mathbf{T}, \delta, \mathbf{y}, \hat{\theta}_y, \hat{\theta}_b) \;\propto\; & \prod_{i=1}^{n} p(T_i, \delta_i \mid \mathbf{b}_i; \theta_s) \\ & \times \prod_{j=1}^{N_i} \prod_{k=1}^{K} p\Big(y_{ik}(t_{ij}) \mid \mathbf{b}_i; \hat{\theta}_y\Big)\, p(\mathbf{b}_i; \hat{\theta}_b)\, p(\theta_s). \end{aligned} \tag{4.3}$$

In this formulation, the random effects $\mathbf{b}_i$ are treated as individual-level parameters with informative prior distributions $p(\mathbf{b}_i; \hat{\theta}_b)$ obtained from the first

stage, where $\hat{\theta}_b = \hat{\mathbf{D}}$. This reduces the model hierarchy by removing the integration over random effects, thereby improving convergence speed and computational efficiency compared to a full joint specification. Moreover, incorporating the longitudinal likelihood function in the second stage acts as a bias correction mechanism, allowing part of the longitudinal uncertainty to be reflected in the survival component. By specifying informative prior distributions based on the first-stage results, the C2S approach allows the random effects to be re-estimated as random variables within the survival submodel rather than fixed covariates. This strategy effectively accounts for the inherent variability of the latent trajectories during the second stage, ensuring that the estimated survival parameters exhibit a posterior dispersion comparable to the fully joint model.

Overall, it turns out that the corrected two-stage approach maintains most of the accuracy of full joint modelling while substantially reducing the computational burden and improving stability in complex settings.

4.3.1.3 Slope-Corrected Two-Stage (SC2S) Approach

The C2S approach proposed by Alvares and Leiva-Yamaguchi (2023) reduces bias in the survival parameters and random effects compared with the standard two-stage (S2S) method. However, in its first step, the C2S approach still ignores the dependence between the longitudinal measurements and the event process. The longitudinal submodel is estimated without accounting for subject dropout due to the occurrence of the event, although this information may be informative and should be taken into account.

The occurrence of the event may strongly affect the observed longitudinal trajectories: subjects at higher risk tend to experience the event earlier and therefore contribute fewer measurements. By fitting the longitudinal model separately in Step 1, two-stage approaches fail to account for this selection mechanism and may lead to biased estimation of the longitudinal time trend (Tsiatis and Davidian, 2004).

This phenomenon was highlighted by Mauff et al. (2020), who showed that two-stage and fully joint (JS) estimators differ substantially at the random-effects level, particularly for the random slopes, whereas the random intercepts show much smaller differences. This suggests that time-related random effects are particularly prone to bias under the two-stage approach. Since the population-level slope parameters (denoted $(\beta_2^{(p)})_{p=1}^{P}$ in (4.6)) reflect the average evolution across individual trajectories, their estimation may be affected by bias in the random slopes.

As discussed in Section 4.3.1.2, the C2S approach partially corrects the bias in the random effects during Step 2 by re-estimating them jointly with the survival parameters. However, the other longitudinal parameters are kept fixed at their Step 1 estimates. In particular, the potential bias affecting the slope parameters $(\beta_2^{(p)})_{p=1}^{P}$ is not corrected.

To address this limitation, we propose an extension of the C2S method, referred to as the Slope-Corrected Two-Stage (SC2S) approach, which allows the longitudinal fixed (population-level) slopes $(\beta_2^{(p)})_{p=1}^{P}$ to be jointly updated in the second stage, together with the random effects and survival parameters. By re-estimating the slope parameters under the joint likelihood instead of relying exclusively on the longitudinal submodel, the SC2S approach incorporates information from both the longitudinal and survival processes and allows correction of the bias introduced at the first stage, when these parameters were estimated separately.

The posterior distribution in (4.3) is therefore modified as follows:

$$\begin{aligned} p(\theta_s, \theta_{y,t}, \mathbf{b} \mid \mathbf{T}, \delta, \mathbf{y}, \hat{\theta}_{y,0}, \hat{\theta}_b) \;\propto\; & \prod_{i=1}^{n} p(T_i, \delta_i \mid \mathbf{b}_i; \theta_s) \\ & \times \prod_{j=1}^{N_i} \prod_{k=1}^{K} p\Big(y_{ik}(t_{ij}) \mid \mathbf{b}_i, \theta_{y,t}; \hat{\theta}_{y,0}\Big) \\ & \times p(\mathbf{b}_i; \hat{\theta}_b)\, p(\theta_s)\, p(\theta_{y,t}) \end{aligned} \tag{4.4}$$

where $\theta_y = (\theta_{y,0}, \theta_{y,t})$, with $\theta_{y,t} = (\beta_2^{(p)})_{p=1}^{P}$ denoting the longitudinal population-level slope parameters updated in the second stage, and $\theta_{y,0}$ collecting the remaining longitudinal parameters, which are kept fixed at their first-stage estimates $\hat{\theta}_{y,0}$.

4.3.2 Prior distributions

We specify weakly informative prior distributions for the full vector of model parameters θ (Papageorgiou et al., 2019; Alsefri et al., 2020). These priors are used consistently across all estimation strategies (JM, SC2S, C2S and S2S) to ensure comparability of inference.

Independent normal priors are assigned to the regression coefficients $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\alpha}$. All free discrimination parameters $\boldsymbol{a}$ are assigned independent normal priors. Several strategies can be used to impose order constraints on

the thresholds. Here, we reparameterize them through their successive first-order differences and place weakly informative half-normal priors $\mathcal{N}^+(0, 10)$ on these differences, which ensures that the thresholds remain ordered.

Rather than assigning a prior directly to the random-effects covariance matrix $\mathbf{D}$, we adopt a decomposition into marginal variances and a correlation matrix. Two alternative specifications can be considered. First, the variances of the random effects are assigned inverse-gamma $\mathrm{IG}(0.01, 0.01)$ priors, while the corresponding correlation parameters follow uniform distributions on $[-1, 1]$. An alternative specification consists of using an LKJ (Lewandowski –Kurowicka–Joe) prior (Lewandowski et al., 2009) for the correlation matrix $\mathbf{R}$, which shrinks correlations toward zero unless strongly supported by the data, together with weakly informative priors (e.g., half-normal) on the standard deviations ($\boldsymbol{\sigma}_b$). The covariance matrix is then parameterized as $\mathbf{D} = \mathrm{diag}(\boldsymbol{\sigma}_b)\,\mathbf{R}\,\mathrm{diag}(\boldsymbol{\sigma}_b)$. These prior choices are standard in the literature and have been shown to provide stable and robust estimation of covariance structures.

The flexibility of the log-baseline hazard, expressed as a linear combination of multiple B-spline basis functions, is regularized using a Gaussian Markov field prior on $\boldsymbol{\gamma}_{h_0}$ (Eilers and Marx, 1996; Lambert and Eilers, 2005):

$$\begin{aligned} p(\boldsymbol{\gamma}_{h_0} \mid \tau) &\propto \tau^{\rho(\mathbf{K})/2} \exp\Big(-\frac{\tau}{2}\,\boldsymbol{\gamma}_{h_0}^\top \mathbf{K} \boldsymbol{\gamma}_{h_0}\Big), \\ \tau &\sim \mathrm{Gamma}(\mathfrak{a}, \mathfrak{b}) \end{aligned} \tag{4.5}$$

where τ is a roughness penalty parameter and $\mathbf{K} = \boldsymbol{\Delta}_r^\top \boldsymbol{\Delta}_r$ is the penalty matrix of rank $\rho(\mathbf{K}) = U - r$ associated with the baseline hazard function h_0, where $\boldsymbol{\Delta}_r$ denotes the rth difference penalty matrix (typically $r = 2$). The amount of smoothness is controlled by τ, with $\mathfrak{a} = 1$ and $\mathfrak{b}$ set to a small value (e.g., 0.005) (Lang and Brezger, 2004).

4.4 Simulation study

We conducted a simulation study to assess the performance of the corrected two-stage approaches, considering both estimation accuracy and computational cost, within the framework of the multivariate latent trait joint model defined in (4.1). This simulation study had two primary aims. First, we evaluated to what extent the corrected two-stage approaches (C2S and SC2S) improve estimation accuracy relative to the standard two-stage approach (S2S), while preserving the computational efficiency of S2S. Second, we evaluated the estimation quality of the SC2S approach by analysing the bias, RMSE, and

coverage probabilities for all parameters of the joint model.

4.4.1 Simulation design

We generated 500 datasets, each including $n = 500$ subjects with up to seven scheduled measurement times. We considered $K = 10$ ordinal outcomes, each with $L_k = 5$ categories. The maximum follow-up time was set to 6. Longitudinal ordinal data were generated from the latent process (see Section 4.2.1)

$$\eta_i^{(p)}(t) = \beta_0^{(p)} + \beta_1^{(p)} x_i + \beta_2^{(p)} t + b_{i0}^{(p)} + b_{i1}^{(p)} t, \qquad p = 1, 2. \tag{4.6}$$

and the event process (see Section 4.2.2) followed

$$h_i(t) = h_0(t) \exp\left\{ x_i \gamma + \sum_{p=1}^{2} \left(\alpha_0^{(p)} b_{i0}^{(p)} + \alpha_1^{(p)} b_{i1}^{(p)} \right) \right\}.$$

We refer to $\alpha_0^{(1)}$ and $\alpha_0^{(2)}$ as the random-intercept association parameters, and to $\alpha_1^{(1)}$ and $\alpha_1^{(2)}$ as the random-slope association parameters. The binary covariate x_i was simulated from Bernoulli(0.5) to mimic a balanced treatment assignment. Regression coefficients were set to

$$\boldsymbol{\beta}^{(1)} = (0.5, -1, 0.75)^\top, \qquad \boldsymbol{\beta}^{(2)} = (1, -0.5, 0.75)^\top, \qquad \gamma = 0.5.$$

Random effects $\mathbf{b}_i = (b_{i0}^{(1)}, b_{i1}^{(1)}, b_{i0}^{(2)}, b_{i1}^{(2)})^\top$ were drawn from a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{D})$, where the covariance matrix $\mathbf{D}$ is defined element-wise by

$$D_{rs} = \rho_{rs}\, \sigma_r\, \sigma_s, \qquad r, s = 1, \ldots, 4,$$

with $\sigma_r^2 = 1$ for all r, and $\rho_{rs} = \rho_{sr}$ denoting the correlation between the rth and sth components of $\mathbf{b}_i$. Specifically, we set $\rho_{12} = 0.30$, $\rho_{13} = 0.20$, $\rho_{14} = 0.10$, $\rho_{23} = 0.10$, $\rho_{24} = 0.20$, and $\rho_{34} = 0.30$.

For identifiability, we set $d_{11} = d_{21} = 0$, $a_1^{(1)} = a_2^{(2)} = 1$, and $a_1^{(2)} = 0$. Values of the discrimination and threshold parameters are reported in Table 4.5 in Appendix C2. Across the 500 simulated datasets, the censoring rate was 10.8%, with a median event time of 2.28 (interquartile interval: 1.09–4.11). Each subject contributed on average 3.27 longitudinal measurements.

In addition, we considered an alternative simulation scenario based on a current-value association structure,

$$m(t, \mathbf{b}_i) = \left(\eta^{(1)}(t), \eta^{(2)}(t) \right).$$

Parameter values for the event and latent processes under this scenario are reported in Table 4.7 (Appendix C3). The discrimination and threshold parameters were kept identical to those used in the random-effects association structure.

MCMC samples for the different models were generated using Stan via the R interface `cmdstanr` (version 2.36.0), relying on the No-U-Turn Sampler (NUTS), an adaptive Hamiltonian Monte Carlo (HMC) algorithm, in Stage 1, and using MCMC algorithms implemented in R for Stage 2. To accelerate computation during Stage 1, the evaluation of the longitudinal likelihood was parallelised over three CPU cores using Stan's `reduce_sum` procedure. Note that all two-stage approaches share the same first-stage estimation procedure.

4.4.2 Simulation results

Table 4.1 summarises the computation time for each method. All two-stage methods substantially reduced runtime relative to the fully joint model (JS), with S2S being the fastest as expected.

Table 4.1: Computation time (in minutes) per dataset

	Mean	Median	Central 95% interval
JS	21.6	21.1	15.4 – 31.8
SC2S	12.3	12.0	8.8 – 18.5
C2S	11.8	11.5	8.3 – 18.0
S2S	11.3	11.0	7.8 – 17.5

Figure 4.1 presents the estimates of the survival coefficients and the association parameters. As anticipated, S2S shows larger bias, particularly for the random-slope association parameters ($\alpha_1^{(1)}$ and $\alpha_1^{(2)}$), whereas bias for the random-intercept association parameters is relatively small. This is consistent with the fact that random intercepts are less influenced by the dropout mechanism. By re-estimating the random effects in the second stage, C2S partially reduces this bias. Our proposed SC2S method provides the largest improvement: for slope-association parameters, SC2S biases were 0.02 and -0.01, compared with -0.09 and -0.11 under S2S.

To examine the accuracy of the uncertainty estimation, Figure 4.2 examines the distribution of the posterior standard deviations across all simulations. As expected, the standard deviation from the S2S estimators is smaller than the competitors. This happens because this approach is structurally unable to propagate first-stage sampling error. However, the bias-corrected methods (C2S and SC2S) provide more realistic measures of uncertainty. It can be ob-

served that the posterior dispersion obtained with our SC2S approach is closer to that of the fully joint specification (JS) than the S2S approach. This confirms that the SC2S strategy effectively propagates longitudinal uncertainty into the survival submodel, resulting in more reliable measures of precision. To conclude, across all survival-related parameters, SC2S approaches to the performance of the fully joint specification, while retaining a computationally efficient two-stage estimation structure.

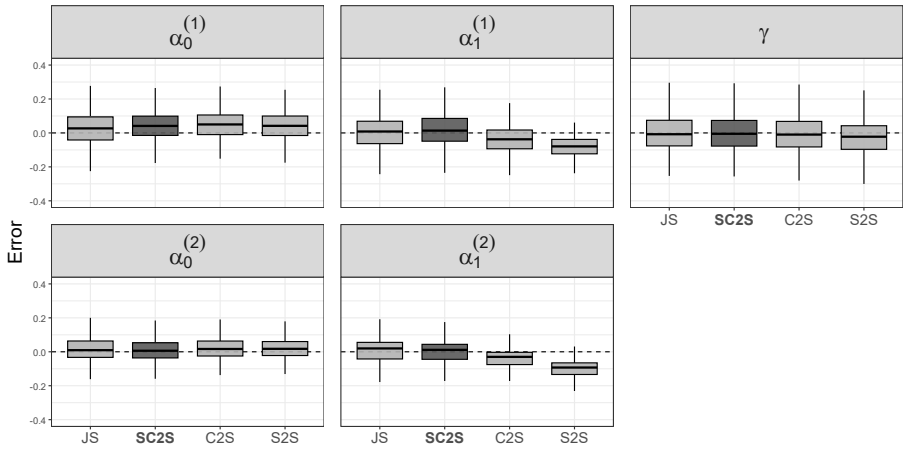

Figure 4.1: Estimation errors for the survival parameters

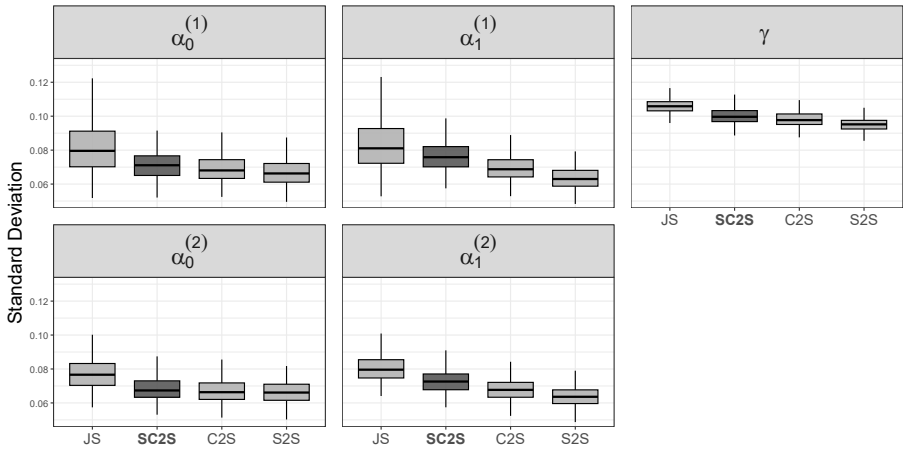

Figure 4.2: Posterior standard deviations for the survival parameters

Figure 4.3 focuses on the longitudinal parameters $\beta_2^{(1)}$ and $\beta_2^{(2)}$, and shows that they are affected by informative dropout under the S2S and C2S approaches. When the joint likelihood is not accounted for in the first stage,

substantial bias arises in these slope fixed effects. Because neither method re-estimates them in the second stage, both S2S and C2S inherit this bias. In contrast, the SC2S approach explicitly re-estimates $(\beta_2^{(p)})_{p=1}^2$ in the second stage using the joint likelihood, leading to a substantial reduction in bias. The estimation of other longitudinal parameters (not shown here), however, does not exhibit notable first-stage bias for any of the methods.

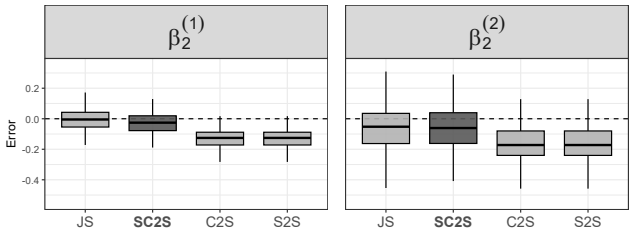

Figure 4.3: Estimation errors of the fixed time effect parameters

To further assess the proximity of the proposed two-stage approaches to the full joint specification (JS), we computed Wasserstein distances between the posterior distributions of the model parameters obtained under each method. This distance provides a natural way to compare entire posterior distributions rather than point estimates, thereby capturing differences in both location and dispersion (Villani, 2009). For each model parameter, it can be interpreted as the minimal cost required to transform one posterior distribution into another. For each simulated dataset, the Wasserstein distance was computed for each parameter by comparing the posterior distributions obtained under JS and each competing approach, and then summarized across parameters. The results are reported in Table 4.2, where distances are aggregated by parameter family, namely the longitudinal fixed effects $\{\beta_j^{(p)}\}$, the survival parameter γ, and the association parameters $\{\alpha_q^{(p)}\}$.

Table 4.2: Mean Wasserstein distances to JS by parameter family across 500 simulated datasets.

Family	JS–SC2S	JS–C2S	JS–S2S
Beta	0.0555	0.0831	0.0831
Gamma	0.0172	0.0198	0.0316
Association	0.0409	0.0556	0.0830

Overall, as expected, the SC2S approach produces posterior distributions that are closest to those obtained under the full joint model, while the S2S approach exhibits the largest distances. The most substantial improvement is

observed for the association parameters $\{\alpha_q^{(p)}\}$, with a reduction of approximately 51% in Wasserstein distance when moving from S2S to SC2S. A more detailed parameter-by-parameter comparison is provided in Table 4.6 (Appendix C2). As anticipated, within the association family, the largest gains are observed for the slope-related parameters (i.e., $\alpha_1^{(p)}$), which are known to be particularly sensitive to informative dropout. Finally, threshold and discrimination parameters are not reported here, as their estimation is identical across the two-stage approaches.

We also examine the asymptotic behaviour of the estimators by studying the evolution of the bias as the sample size increases (from $n = 250$ to $n = 1000$) for selected parameters. The results, displayed in Figure 4.5 (Appendix C2), show a clear reduction in both bias and dispersion as n increases for both the JS and SC2S approaches. Moreover, the SC2S method closely follows the behaviour of the joint model.

The performance measures obtained with the SC2S approach are reported in Table 4.3 and in Table 4.5 in Appendix C2. Across all classes of parameters, we observe very small estimation biases and coverage probabilities close to the nominal level using SC2S. In particular, the good performances obtained for the discrimination and threshold parameters confirm that the model successfully differentiates items and correctly identifies the ordering of response categories. For the longitudinal parameters, the fixed effects of both latent dimensions are accurately recovered. As discussed, the estimation accuracy for the coefficients $(\beta_2^{(p)})_{p=1}^2$ is slightly lower with SC2S than with the JS, but the bias is largely reduced compared with the classical two-stage approach. Finally, in the survival submodel, the association parameters are well recovered, demonstrating that the method effectively captures the distinct random effects associated with the two latent traits.

The corresponding results under the current-value association structure are reported in Tables 4.7 and 4.8 (Appendix C3). The SC2S approach performs consistently in the current-value association setting, exhibiting very small estimation biases and coverage probabilities close to the nominal level. In this setting, the computation time is 12.9 (8.6–26.9) minutes with the SC2S approach, compared with 30.4 (21.3 – 49.5) with the JS approach.

Taken together, these findings demonstrate that SC2S achieves a favourable balance between accuracy and computational efficiency, closely approximating the fully joint estimator while reducing runtime.

Table 4.3: Simulation results for $S = 500$ replications of sample of size $n = 500$ with the SC2S approach : bias, root mean square error (RMSE) and effective coverage (COV) of 95% credible intervals for the survival and latent processes parameters.

Parameter	True	Bias	RMSE	COV
$\beta_0^{(1)}$	0.50	0.01	0.10	96%
$\beta_0^{(2)}$	1.00	-0.03	0.14	98%
$\beta_1^{(1)}$	-1.00	0.00	0.12	94%
$\beta_1^{(2)}$	-0.50	0.06	0.20	99%
$\beta_2^{(1)}$	0.75	-0.03	0.08	81%
$\beta_2^{(2)}$	0.75	-0.06	0.16	75%
γ	0.50	-0.01	0.12	91%
$\alpha_0^{(1)}$	0.30	0.03	0.10	88%
$\alpha_1^{(1)}$	0.30	0.02	0.10	88%
$\alpha_0^{(2)}$	0.30	0.01	0.07	96%
$\alpha_1^{(2)}$	0.30	-0.01	0.07	93%

4.5 Application to glioblastoma data

The dataset used in this study originates from a phase III clinical trial involving patients with first progression of glioblastoma. This trial was conducted by the European Organisation for Research and Treatment of Cancer (EORTC-26101 trial) to determine whether the combination of lomustine and bevacizumab (combination group) improved overall survival compared to lomustine alone (monotherapy group). Details of this trial can be found in ClinicalTrials.gov (2021) (Identifier NCT01290939). While the primary endpoint of the trial is overall survival, these patients have a poor prognosis, making it crucial to evaluate the treatment's impact on their well-being.

To this end, quality-of-life data were collected longitudinally using the EORTC Quality of Life Questionnaires QLQ-C30 and QLQ-BN20. The latter module specifically assesses the effects of brain tumours and their treatment on symptoms, functioning, and health-related quality of life (HRQoL), whereas the former was developed for use in general cancer populations (Fayers et al., 2001).

Specifically, the QLQ-BN20 is composed of 20 ordinal items measuring different dimensions of quality of life including future uncertainty (four items), visual disorder (three items), motor dysfunction (three items), and communi-

cation deficit (three items). Seven additional single items assess headaches, seizures, drowsiness, hair loss, itchy skin, weakness of legs, and bladder control. Patients self-reported their responses at baseline and every 12 weeks thereafter until disease progression, allowing the monitoring of changes in HRQoL over time.

Trial results were previously published (Wick et al., 2017), and did not demonstrate an overall survival advantage of lomustine plus bevacizumab over lomustine alone in patients with progressive glioblastoma, despite a prolonged progression-free survival (PFS). At that time, HRQoL was analysed using a sum-score approach and did not identify a significant treatment effect.

To demonstrate our methodology, we focus on the motor dysfunction (MD) and communication deficit (CD) scales, as these were identified as the primary domains of interest in Wick et al. (2017). These scales are defined as follows:

Motor dysfunction

Item 1. *Did you have weakness on one side of your body?*

Item 2. *Did you have trouble with your coordination?*

Item 3. *Did you feel unsteady on your feet?*

Communication deficit

Item 4. *Did you have trouble finding the right words to express yourself?*

Item 5. *Did you have difficulty speaking?*

Item 6. *Did you have trouble communicating your thoughts?*

Response categories:

1 = “Not at all” 2 = “A little” 3 = “Quite a bit” 4 = “Very much”

Each item has a response range from 1 ("Not at all") to 4 ("Very much"), where higher responses indicate worse dysfunction or deficit. Appendix C4 presents an exploratory visualization of the raw longitudinal ordinal responses for the motor dysfunction (Items 1–3) and communication deficit (Items 4–6) scales. We impose identifiability constraints on one item from each dimension, that is, Item 1 and Item 4, by letting $d_{11} = d_{41} = 0$, diagonal elements of matrix $(a_1, a_4)^\top$ equal to 1 and off-diagonal elements equal to 0. These items were selected based on the exploratory two-dimensional GRM analysis presented in Appendix C5.

The dataset analysed includes $n = 423$ patients enrolled across multiple institutions in Europe. Of these, 145 were randomised to receive monotherapy

treatment, while the rest were randomised to receive a combination of lomustine and bevacizumab. Follow-up times for the longitudinal HRQoL data ranged from 1 to 96 weeks. The median overall survival was 40.4 weeks, and the median progression-free survival was 13.6 weeks. We consider progression-free survival as the event of interest for the survival component. Each patient completed the HRQoL questionnaire between 1 and 9 times during follow-up.

We consider the following joint model in our analysis:

$$\begin{cases} h_i(t) = h_0(t) \exp\left\{w_i\gamma + \sum_{p=1}^{2}\left(\alpha_0^{(p)} b_{i0}^{(p)} + \alpha_1^{(p)} b_{i1}^{(p)}\right)\right\}, \\ \eta_i^{(1)}(t) = \beta_0^{(1)} + \beta_1^{(1)} x_i + \beta_2^{(1)} t + b_{i0}^{(1)} + b_{i1}^{(1)} t, \\ \eta_i^{(2)}(t) = \beta_0^{(2)} + \beta_1^{(2)} x_i + \beta_2^{(2)} t + b_{i0}^{(2)} + b_{i1}^{(2)} t, \\ \text{logit}\left\{\Pr\left(y_{ik}(t) \le l \mid \boldsymbol{\eta}_i(t)\right)\right\} = d_{kl} - \sum_{p=1}^{2} a_k^{(p)} \eta_i^{(p)}(t). \end{cases}$$

where $\eta_i^{(1)}(t)$ and $\eta_i^{(2)}(t)$ represent the latent motor dysfunction and communication deficit of patient i at time t. Here, $l = 1, \ldots, 4$ indicates response categories and $k = 1, \ldots, 6$ the six items. Higher latent scores correspond to worse functioning. We include a random intercept and a random slope for each latent dimension, resulting in $\mathbf{b}_i = (b_{i0}^{(1)}, b_{i1}^{(1)}, b_{i0}^{(2)}, b_{i1}^{(2)})^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$, with $D_{rr} = \sigma_r^2$ and $D_{rs} = \rho_{rs}\sigma_r\sigma_s$ for $r, s = 1, \ldots, 4$, $r < s$. In this application, the same treatment variable is included in both $\mathbf{w}_i$ and $\mathbf{x}_i$.

We use the slope-corrected two-stage (SC2S) approach to estimate this joint model, the computation took about nine minutes. Trace plots revealed no discernible trends in the MCMC chains, indicating adequate convergence. Furthermore, the Gelman–Rubin diagnostic (Gelman and Rubin, 1992) confirmed that the potential scale reduction factor (R) for all parameters remained below 1.1.

4.5.1 Results overview

Estimated parameters for the latent and survival processes are reported in Table 4.4. The results for the GRM parameters and for the covariance matrix are presented in Tables 4.10 and 4.11 in Appendix C6.

First, on the longitudinal dimensions, Table 4.4 reveals that treatment has a significant effect on the level of motor dysfunction. The positive coefficient ($\beta_1^{(1)} = 0.606$), with the monotherapy group as reference and given that higher latent scores reflect worse functioning, indicates that patients in the combination group have, on average, worse motor functioning. By contrast,

Table 4.4: Estimation results for latent processes and survival submodel parameters : Posterior means and 95% credible intervals (CI) are reported.

Parameter	Est.	CI 95%
Motor dysfunction		
$\beta_0^{(1)}$	-1.530	[-1.983; -1.137]
$\beta_1^{(1)}$	0.606	[0.187; 1.073]
$\beta_2^{(1)}$	0.010	[-0.011; 0.031]
Communication deficit		
$\beta_0^{(2)}$	0.925	[0.131; 1.822]
$\beta_1^{(2)}$	-0.425	[-1.379; 0.511]
$\beta_2^{(2)}$	0.060	[0.024; 0.098]
Progression-free survival		
γ	-1.411	[-2.123; -0.882]
$\alpha_0^{(1)}$	-0.286	[-0.759; 0.075]
$\alpha_1^{(1)}$	-3.760	[-24.79; 21.09]
$\alpha_0^{(2)}$	0.167	[-0.027; 0.396]
$\alpha_1^{(2)}$	18.54	[8.88; 32.80]

treatment did not demonstrate a significant effect on communication deficit. Over time (measured in weeks), communication deficit deteriorated significantly ($\beta_2^{(2)} = 0.060$), whereas the time trend in motor dysfunction was non-significant ($\beta_2^{(1)} = 0.010$).

In the survival submodel, we observed a statistically significant treatment effect on progression-free survival ($\gamma = -1.411$). The association parameters linking individual deviations in the latent processes to the risk of progression highlight communication as the most prognostic dimension. In particular, the association with the random slope of communication deficit is large and significant ($\alpha_1^{(2)} = 18.54$). The magnitude of this estimate should be interpreted in light of the small between-patient variability of the communication slopes (Table 4.11: $\sigma_4 = 0.079$). To improve the interpretability of this effect, we considered standardised association parameters. We obtained $\alpha_{1,\text{std}}^{(2)} = \alpha_1^{(2)} \times \sigma_4 = 1.47$, which represents the log-hazard increase associated with a one standard-deviation increase in the communication random slope. In comparison, the standardised effect for the motor slope was much smaller and non-significant ($\alpha_{1,\text{std}}^{(1)} = \alpha_1^{(1)} \times \sigma_2 = -0.14$). We therefore observed that patients exhibiting faster deterioration in communication abilities tend to have a higher risk of progression. Awareness of this relationship may

encourage clinicians to monitor communication abilities carefully, as an accelerated decline may be indicative of disease progression. Such changes may prompt clinicians to arrange additional assessments (e.g., imaging) or adjust follow-up intensity. Furthermore, supportive interventions may be considered to help patients manage communication difficulties associated with disease progression (e.g., speech therapy).

At the measurement level, the graded response model behaved as expected and clearly recovered the two latent dimensions. It successfully identified the appropriate items loading on each component. Specifically, for all items, the highest and statistically significant discrimination parameters corresponded to their expected latent domain (i.e., motor dimension for Items 1–3 and communication dimension for Items 4–6). In detail, discrimination parameters for items 1, 2, and 3 were concentrated on the motor factor (Component 1), with Item 3 ("unsteady on your feet") showing the highest discrimination ($a_3^{(1)} = 1.656$), while cross-loadings on the communication factor were negligible (e.g., $a_3^{(2)} \approx 0$). For communication items (Component 2), Item 5 ("difficulty speaking") showed the strongest discrimination ($a_5^{(2)} = 1.124$). Also, Item 2 shows a small but significant cross-loading, with a stronger contribution to the motor component.

Figure 4.4 presents the estimated distribution of response probabilities for Items 3 and 5, identified as the most discriminative items for $\eta^{(1)}$ and $\eta^{(2)}$, respectively. It illustrates the temporal evolution of the probabilities that an average patient selects a given response category $l \in 1, \ldots, 4$ at various time points. For Item 3 (motor domain), a clear treatment effect is visible, with consistently worse predicted response distributions under combination therapy. Item 5 (communication domain) shows a pronounced deterioration over time, consistent with the significant positive slope estimated for the communication latent process. Across items, higher values of the latent processes align with increasing probabilities of selecting more severe response categories, as expected under the graded response model. The cumulative hazard curves in the bottom panels show that higher progression risk is associated with higher communication decline (e.g., patients increasingly choosing the "Quite a bit" category for Item 5). A complementary visualization of the estimated response probability distributions for all items under the combination treatment is provided in Appendix C7, where higher response categories and more pronounced changes over time are observed for communication items (Items 4–6) compared with motor items (Items 1–3).

Finally, the random-effects covariance structure indicates meaningful associ-

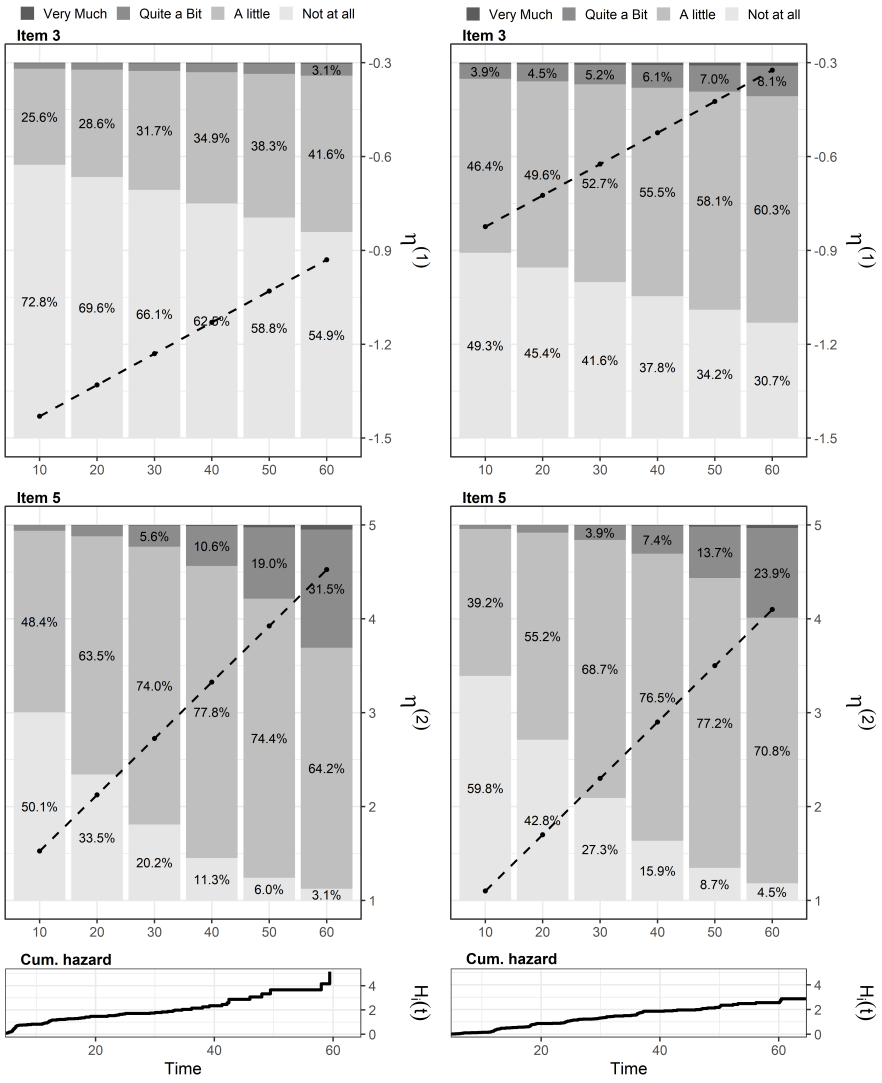

Figure 4.4: Predicted response probabilities over time for Item 3 (top) and Item 5 (middle) under monotherapy (left) and combination therapy (right). Stacked bars represent response probabilities, dashed lines the expected latent trajectories. Bottom panels show cumulative hazard functions.

ations between the two domains, both at baseline and in longitudinal change (Table 4.11). Random intercepts for motor and communication were positively correlated ($\rho_{13} = 0.403$), suggesting that patients with worse baseline motor function tended to have worse baseline communication. Similarly, the random slopes were positively correlated ($\rho_{24} = 0.527$), indicating that patients who deteriorated faster in motor function also tended to deteriorate faster in communication.

4.6 Discussion

In this work, we proposed a slope-corrected two-stage (SC2S) approach for the joint analysis of multivariate ordinal longitudinal data and time-to-event outcomes within a multidimensional latent trait framework. This methodology provides a computationally efficient tool for the analysis of HRQoL data in oncology and other clinical research settings where multidimensional patient-reported outcomes and terminal event are present.

Our method extends the corrected two-stage approach proposed by Alvares and Leiva-Yamaguchi (2023), which uses estimates from the longitudinal submodel to define informative prior distributions for the random effects in the survival submodel. In contrast to their work, we apply this strategy within a substantially more complex modelling framework, involving multivariate ordinal outcomes and multidimensional latent processes. We further extend the C2S by allowing longitudinal slope parameters to be re-estimated in the second stage, thereby correcting the bias introduced by informative dropout.

Through simulation studies, we demonstrated that both SC2S and C2S improve the estimation performance compared with the standard two-stage approach. In particular, SC2S further reduces bias in survival association parameters and in longitudinal slope parameters. Importantly, its performance approaches that of fully joint Bayesian estimation (JS), while offering considerable reductions in computation time. We also evaluated the quality of uncertainty quantification through the posterior standard deviations. The results show that the standard two-stage approach tends to underestimate variability, as expected. In contrast, both C2S and SC2S provide more realistic measures of uncertainty. This highlights the importance of accounting for the variability of the latent trajectories when estimating survival parameters. To further assess the proximity to the fully joint model, we compared posterior distributions using the Wasserstein distance, which captures differences in both location and dispersion. The results show that SC2S produces posterior distributions closest to those obtained under JS. Finally, we note that we did not rely on predictive criteria such as the WAIC for model comparison. Since

all approaches are based on the same underlying joint model, differences arise from the estimation strategy rather than from the model specification itself. In this context, such criteria are not particularly informative for discriminating between approaches. This explains why we focused on parameter estimation accuracy and on the similarity of posterior distributions with the fully joint model, which are more relevant criteria for evaluating two-stage estimation procedures.

Beyond these performance results, some practical considerations regarding model specification deserve attention. In this approach, accurate estimation of the random-effects covariance matrix $\mathbf{D}$ is crucial, as it directly determines the prior distribution of the random effects used in the second stage, and its misspecification may propagate through the model. Careful prior specification for $\mathbf{D}$ is therefore essential. Indeed, the impact of the prior on $\mathbf{D}$ depends on the amount of longitudinal information available. With few observations per subject, the prior can have a substantial influence, making robust prior specification especially important. Both specifications described in Section 4.3.2 are valid. However, in practice, the LKJ-based parameterization is generally preferred, as it ensures a proper correlation structure and provides more stable inference, especially in higher dimensions.

The methodology was applied to longitudinal HRQoL data from a phase III glioblastoma trial. The proposed joint model identified clinically relevant treatment effects and associations between latent functional dimensions and progression-free survival. In particular, the survival analysis highlighted communication decline as a prognostic marker: patients whose communication abilities deteriorate more rapidly are at increased risk of progression, as reflected by a significant association with the communication random slope. Moreover, the graded response model accurately recovered the expected latent structure of the HRQoL items. The model also captured meaningful correlation patterns between motor and communication deficits, indicating that patients starting worse or declining faster in one domain tended to do so in the other as well. These findings show the ability of the proposed approach to uncover clinically relevant patterns in multidimensional patient-reported outcomes and to detect subtle changes that may not be detected by univariate or traditional score-based analyses.

Finally, an advantage of the corrected two-stage approaches is that they can be easily generalised to many settings. They can accommodate a wide range of longitudinal models (including flexible trajectories and different outcome types) as well as more complex survival models such as competing-risks or multistate models. Elaborate association structures could also be considered.

4.7 Appendix C

C1 Distributional assumptions

$$
\begin{aligned}
p(T_i, \delta_i \mid \mathbf{b}_i; \theta_s) &= \big[h_i(T_i \mid \mathbf{b}_i; \theta_s)\big]^{\delta_i} S_i(T_i \mid \mathbf{b}_i; \theta_s) \\
&= \Big[h_0(T_i)\exp\{\boldsymbol{\gamma}^\top \mathbf{w}_i + \alpha\, m(T_i, \mathbf{b}_i)\}\Big]^{\delta_i} \\
&\quad \times \exp\left(-\int_0^{T_i} h_0(s)\exp\{\boldsymbol{\gamma}^\top \mathbf{w}_i + \alpha\, m(s, \mathbf{b}_i)\}\, ds\right) \\
&= \left[\exp\left\{\sum_{u=1}^{U} \gamma_{h_0,u} B_u(T_i) + \boldsymbol{\gamma}^\top \mathbf{w}_i + \alpha\, m(T_i, \mathbf{b}_i)\right\}\right]^{\delta_i} \\
&\quad \times \exp\left[-\exp\{\boldsymbol{\gamma}^\top \mathbf{w}_i\}\int_0^{T_i} \exp\left\{\sum_{u=1}^{U} \gamma_{h_0,u} B_u(s) + \alpha\, m(s, \mathbf{b}_i)\right\} ds\right]
\end{aligned}
$$

$$
\begin{aligned}
p\big(y_{ik}(t) \mid \mathbf{b}_i; \theta_y\big) &= \prod_{l=1}^{L_k} \Pr\big(y_{ik}(t) = l \mid \boldsymbol{\eta}_i(t)\big)^{\mathbb{1}\{y_{ik}(t)=l\}} \\
&= \Big\{\operatorname{expit}\big(d_{k1} - \boldsymbol{\alpha}_k^\top \boldsymbol{\eta}_i(t)\big)\Big\}^{\mathbb{1}\{y_{ik}(t)=1\}} \\
&\quad \times \prod_{l=2}^{L_k-1} \Big\{\operatorname{expit}\big(d_{kl} - \boldsymbol{\alpha}_k^\top \boldsymbol{\eta}_i(t)\big) - \operatorname{expit}\big(d_{k,l-1} - \boldsymbol{\alpha}_k^\top \boldsymbol{\eta}_i(t)\big)\Big\}^{\mathbb{1}\{y_{ik}(t)=l\}} \\
&\quad \times \Big\{1 - \operatorname{expit}\big(d_{k,L_k-1} - \boldsymbol{\alpha}_k^\top \boldsymbol{\eta}_i(t)\big)\Big\}^{\mathbb{1}\{y_{ik}(t)=L_k\}}
\end{aligned}
$$

where $\operatorname{expit}(x) = \frac{1}{1+e^{-x}}$.

$$
p(\mathbf{b}_i; \theta_b) = (2\pi)^{-q/2} \det(\mathbf{D})^{-1/2} \exp\big(-\tfrac{1}{2}\mathbf{b}_i^\top \mathbf{D}^{-1} \mathbf{b}_i\big)
$$

where q denotes the dimension of the random effect vector.

C2 Simulation results for the random-effects (RE) association

Table 4.5: Bias, RMSE, and 95% coverage for discrimination (top) and threshold (bottom) parameters under the SC2S approach with RE association

	True	Bias	RMSE	COV		True	Bias	RMSE	COV
$a_1^{(1)}$	1.00	/	/	/	$a_1^{(2)}$	0.00	/	/	/
$a_2^{(1)}$	0.10	0.07	0.18	98%	$a_2^{(2)}$	1.00	/	/	/
$a_3^{(1)}$	1.50	0.05	0.13	95%	$a_3^{(2)}$	0.40	0.01	0.05	95%
$a_4^{(1)}$	2.00	0.05	0.13	95%	$a_4^{(2)}$	0.10	-0.00	0.06	95%
$a_5^{(1)}$	0.10	0.11	0.29	99%	$a_5^{(2)}$	1.60	0.02	0.10	95%
$a_6^{(1)}$	1.20	0.04	0.09	95%	$a_6^{(2)}$	0.20	0.00	0.04	95%
$a_7^{(1)}$	1.60	0.05	0.13	96%	$a_7^{(2)}$	0.30	0.00	0.05	95%
$a_8^{(1)}$	0.20	0.08	0.22	99%	$a_8^{(2)}$	1.20	0.01	0.07	92%
$a_9^{(1)}$	0.30	0.07	0.19	98%	$a_9^{(2)}$	1.00	0.01	0.06	95%
$a_{10}^{(1)}$	0.40	0.11	0.28	98%	$a_{10}^{(2)}$	1.50	0.03	0.09	92%

	True	Bias	RMSE	COV		True	Bias	RMSE	COV
$d_{1,1}$	0.00	/	/	/	$d_{1,2}$	1.00	0.00	0.07	95%
$d_{2,1}$	0.00	/	/	/	$d_{2,2}$	1.00	0.01	0.06	96%
$d_{3,1}$	1.00	0.03	0.14	96%	$d_{3,2}$	2.00	0.04	0.16	95%
$d_{4,1}$	2.00	0.04	0.19	95%	$d_{4,2}$	3.00	0.05	0.21	97%
$d_{5,1}$	0.80	0.01	0.15	97%	$d_{5,2}$	1.80	0.02	0.16	96%
$d_{6,1}$	0.50	0.02	0.13	90%	$d_{6,2}$	1.50	0.03	0.14	94%
$d_{7,1}$	0.30	0.02	0.14	97%	$d_{7,2}$	1.30	0.02	0.14	97%
$d_{8,1}$	0.60	0.01	0.12	96%	$d_{8,2}$	1.60	0.01	0.13	94%
$d_{9,1}$	1.20	0.02	0.11	95%	$d_{9,2}$	2.20	0.02	0.12	97%
$d_{10,1}$	1.50	0.03	0.16	94%	$d_{10,2}$	2.50	0.04	0.18	94%
$d_{1,3}$	2.50	-0.00	0.10	95%	$d_{1,4}$	5.00	0.02	0.18	95%
$d_{2,3}$	2.70	0.01	0.10	95%	$d_{2,4}$	5.50	0.01	0.19	95%
$d_{3,3}$	3.50	0.05	0.19	94%	$d_{3,4}$	6.00	0.07	0.25	96%
$d_{4,3}$	4.50	0.08	0.25	95%	$d_{4,4}$	7.00	0.13	0.34	95%
$d_{5,3}$	3.30	0.03	0.19	95%	$d_{5,4}$	5.80	0.06	0.27	94%
$d_{6,3}$	3.00	0.06	0.17	91%	$d_{6,4}$	5.50	0.08	0.24	95%
$d_{7,3}$	2.80	0.05	0.16	94%	$d_{7,4}$	5.30	0.07	0.23	95%
$d_{8,3}$	3.10	0.02	0.15	96%	$d_{8,4}$	5.60	0.05	0.23	93%
$d_{9,3}$	3.70	0.03	0.15	96%	$d_{9,4}$	6.20	0.04	0.23	96%
$d_{10,3}$	4.00	0.06	0.22	93%	$d_{10,4}$	6.50	0.09	0.29	94%

Table 4.6: Mean Wasserstein distances to the joint specification (JS) across 500 simulated datasets for selected survival and latent process parameters.

Parameter	JS–SC2S	JS–C2S	JS–S2S
$\beta_0^{(1)}$	0.0085	0.0085	0.0085
$\beta_0^{(2)}$	0.0681	0.0681	0.0681
$\beta_1^{(1)}$	0.0114	0.0114	0.0114
$\beta_1^{(2)}$	0.1396	0.1396	0.1396
$\beta_2^{(1)}$	0.0359	0.1225	0.1225
$\beta_2^{(2)}$	0.0692	0.1487	0.1487
γ	0.0172	0.0198	0.0316
$\alpha_0^{(1)}$	0.0648	0.0673	0.0689
$\alpha_1^{(1)}$	0.0642	0.0836	0.1338
$\alpha_0^{(2)}$	0.0172	0.0209	0.0203
$\alpha_1^{(2)}$	0.0174	0.0507	0.1092

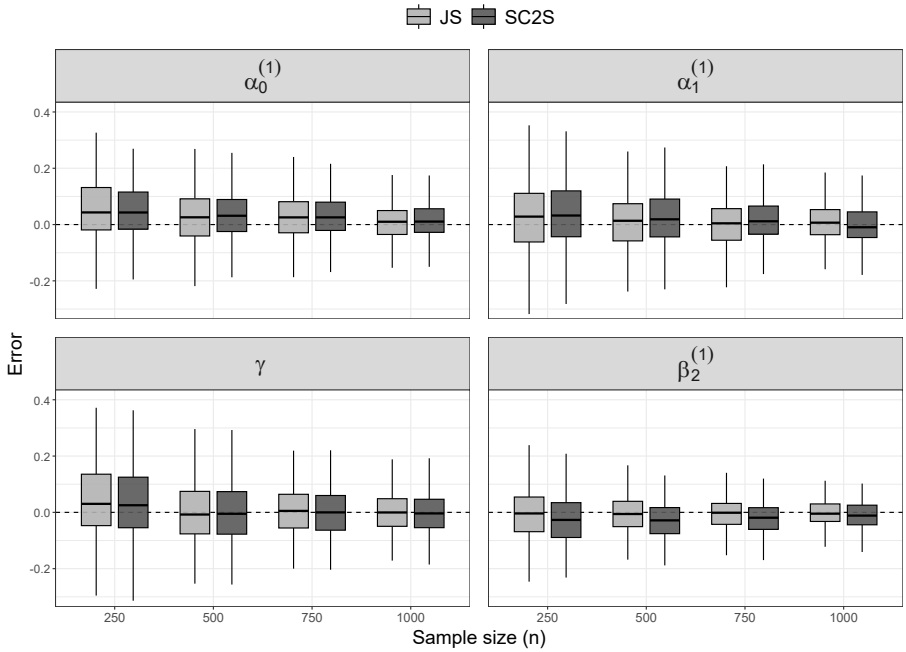

Figure 4.5: Estimation error as a function of the sample size for selected parameters under JS and SC2S

C3 Simulation results for the current-value (CV) association

Table 4.7: Bias, RMSE, and 95% coverage for the survival and latent processes parameters under the SC2S approach with CV association.

Parameter	True	Bias	RMSE	COV
$\beta_0^{(1)}$	-0.50	-0.00	0.08	94%
$\beta_0^{(2)}$	-0.30	0.00	0.09	92%
$\beta_1^{(1)}$	0.10	-0.00	0.08	92%
$\beta_1^{(2)}$	0.15	0.00	0.08	92%
$\beta_2^{(1)}$	0.50	-0.01	0.09	94%
$\beta_2^{(2)}$	0.75	-0.02	0.10	90%
γ	0.30	0.01	0.09	92%
$\alpha^{(1)}$	0.15	-0.05	0.08	89%
$\alpha^{(2)}$	0.10	-0.03	0.07	92%

Table 4.8: Bias, RMSE, and 95% coverage for discrimination (top) and threshold (bottom) parameters under the SC2S approach with CV association

	True	Bias	RMSE	COV		True	Bias	RMSE	COV
$a_1^{(1)}$	0.00	/	/	/	$a_1^{(2)}$	1.00	/	/	/
$a_2^{(1)}$	0.00	/	/	/	$a_2^{(2)}$	1.00	/	/	/
$a_3^{(1)}$	1.00	-0.01	0.13	96%	$a_3^{(2)}$	2.00	-0.00	0.14	98%
$a_4^{(1)}$	2.00	-0.01	0.19	90%	$a_4^{(2)}$	3.00	-0.00	0.21	94%
$a_5^{(1)}$	0.80	0.01	0.14	95%	$a_5^{(2)}$	1.80	0.01	0.15	92%
$a_6^{(1)}$	0.50	-0.01	0.11	94%	$a_6^{(2)}$	1.50	-0.00	0.10	97%
$a_7^{(1)}$	0.30	-0.01	0.13	94%	$a_7^{(2)}$	1.30	-0.01	0.14	92%
$a_8^{(1)}$	0.60	0.02	0.11	95%	$a_8^{(2)}$	1.60	0.02	0.12	95%
$a_9^{(1)}$	1.20	0.01	0.09	96%	$a_9^{(2)}$	2.20	0.01	0.11	97%
$a_{10}^{(1)}$	1.50	0.03	0.15	94%	$a_{10}^{(2)}$	2.50	0.03	0.16	95%

	True	Bias	RMSE	COV		True	Bias	RMSE	COV
$d_{1,1}$	1.00	/	/	/	$d_{1,2}$	0.00	/	/	/
$d_{2,1}$	0.00	/	/	/	$d_{2,2}$	0.00	/	/	/
$d_{3,1}$	1.50	0.08	0.19	95%	$d_{3,2}$	1.00	0.03	0.14	96%
$d_{4,1}$	2.00	0.11	0.26	89%	$d_{4,2}$	2.00	0.04	0.19	95%
$d_{5,1}$	0.10	0.02	0.15	95%	$d_{5,2}$	0.80	0.01	0.15	97%
$d_{6,1}$	1.20	0.06	0.16	89%	$d_{6,2}$	0.50	0.02	0.13	90%
$d_{7,1}$	1.60	0.06	0.20	94%	$d_{7,2}$	0.30	0.02	0.14	97%
$d_{8,1}$	0.20	-0.00	0.13	92%	$d_{8,2}$	0.60	0.01	0.12	96%
$d_{9,1}$	0.30	0.02	0.11	96%	$d_{9,2}$	1.20	0.02	0.11	95%
$d_{10,1}$	0.40	0.02	0.14	97%	$d_{10,2}$	1.50	0.03	0.16	94%
$d_{1,3}$	0.00	/	/	/	$d_{1,4}$	0.00	/	/	/
$d_{2,3}$	1.00	/	/	/	$d_{2,4}$	1.00	/	/	/
$d_{3,3}$	0.40	-0.03	0.11	94%	$d_{3,4}$	0.40	-0.03	0.11	94%
$d_{4,3}$	0.10	-0.05	0.16	92%	$d_{4,4}$	0.10	-0.05	0.16	92%
$d_{5,3}$	1.60	0.01	0.14	94%	$d_{5,4}$	1.60	0.01	0.14	94%
$d_{6,3}$	0.20	-0.02	0.10	95%	$d_{6,4}$	0.20	-0.02	0.10	95%
$d_{7,3}$	0.30	-0.03	0.11	95%	$d_{7,4}$	0.30	-0.03	0.11	95%
$d_{8,3}$	1.20	0.02	0.12	92%	$d_{8,4}$	1.20	0.02	0.12	92%
$d_{9,3}$	1.00	0.00	0.10	95%	$d_{9,4}$	1.00	0.00	0.10	95%
$d_{10,3}$	1.50	0.02	0.14	92%	$d_{10,4}$	1.50	0.02	0.14	92%

C4 Visualization of the raw HRQoL data

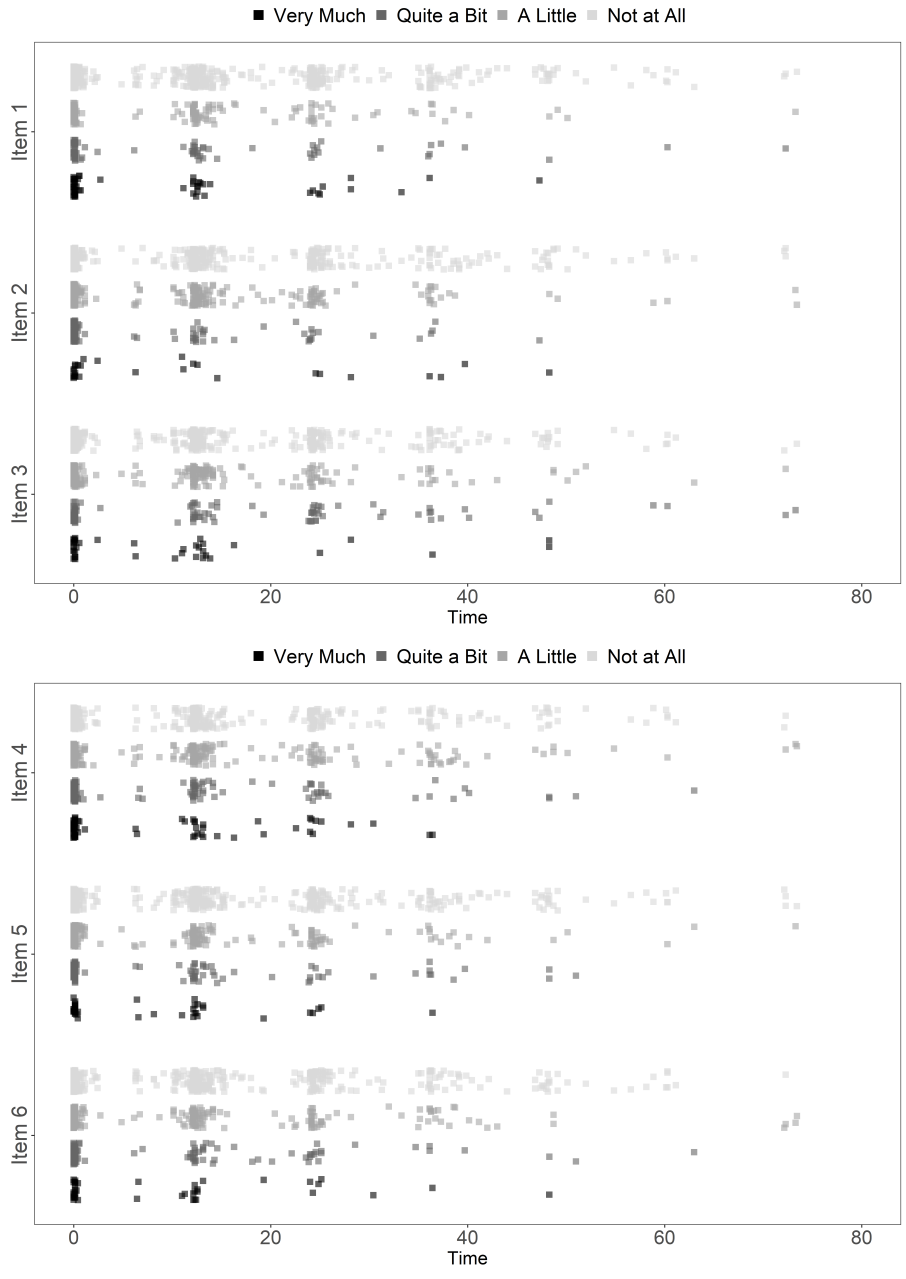

Figure 4.6: Distribution of the raw longitudinal ordinal data for the motor dysfunction (MD, top) and communication deficit (CD, bottom) scales.

C5 Exploratory two-dimensional graded response model

Before specifying the confirmatory multidimensional graded response model, we conducted a preliminary exploratory item factor analysis. The objectives were twofold: (i) to assess whether the six selected items exhibit a two-dimensional latent structure, and (ii) to guide the choice of anchor items for model identification in the confirmatory multidimensional analysis, rather than relying on an arbitrary selection.

We note that objective (i) is not intended to reassess the dimensional validity of the EORTC QLQ-BN20 questionnaire, whose scale structure has been specifically designed and extensively validated to ensure that items load exclusively on their intended domains. Rather, this first step is included primarily as an illustrative example and a consistency check in the present dataset.

In multidimensional IRT models, identifiability is typically achieved by selecting anchor items, namely items loading strongly on a single latent dimension, and fixing their discrimination parameters to define the scale and orientation of each latent trait.

We performed an exploratory two-factor IRT analysis using a graded response formulation. Table 4.9 reports the rotated factor loadings associated with the two latent dimensions (factors F1 and F2), along with the communality (h^2) for each item. Rotated factor loadings reflect the strength of association between each item and the latent dimensions, with higher absolute values indicating greater informativeness with respect to the corresponding factor. The communality (h^2) represents the proportion of an item's variance jointly explained by the latent factors.

The exploratory analysis reveals a clear and well-defined two-factor structure. Items 1, 2, and 3 (motor items) load predominantly on the first factor (F1), which primarily captures the Motor Dysfunction domain. Conversely, Items 4, 5, and 6 (communication items) load strongly on the second factor (F2) corresponding to the Communication Deficit domain. Cross-loadings are negligible for all items, supporting a clear separation between the two latent dimensions.

These results support the use of a two-dimensional latent structure in the confirmatory multidimensional graded response model. Moreover, Item 1 (Motor Dysfunction) and Item 4 (Communication Deficit) exhibit the strongest and most specific loadings on their respective dimensions, making them natural candidates for identification constraints in the multidimensional mea-

surement model.

Table 4.9: Rotated exploratory IRT loadings from a two–factor graded response model.

Item	F1	F2	h^2
1	**0.833**	−0.072	0.643
2	0.753	0.079	0.628
3	0.709	0.038	0.529
4	−0.044	**0.958**	0.881
5	0.005	0.942	0.891
6	0.055	0.896	0.852

C6 Additional results from the application

Table 4.10: Estimation results of GRM submodel parameters: posterior means and 95% credible intervals (CI).

		Est.	CI 95%
Item1	d_{11}	0.000	/
	d_{12}	1.547	[1.304; 1.815]
	d_{13}	2.833	[2.455; 3.247]
Item2	d_{21}	-0.416	[-0.717;-0.153]
	d_{22}	1.993	[1.675; 2.313]
	d_{23}	3.949	[3.448; 4.514]
Item3	d_{31}	-1.406	[-2.122;-0.866]
	d_{32}	1.730	[1.285; 2.173]
	d_{33}	4.185	[3.523; 5.042]
Item4	d_{41}	0.000	/
	d_{42}	4.244	[3.633; 4.964]
	d_{43}	7.259	[6.268; 8.400]
Item5	d_{51}	1.518	[0.988; 2.101]
	d_{52}	5.678	[4.732; 6.898]
	d_{53}	9.376	[7.917; 11.25]
Item6	d_{61}	0.664	[0.264; 1.081]
	d_{62}	4.559	[3.887; 5.332]
	d_{63}	7.618	[6.562; 8.816]

	Est.	CI 95%
Dimension 1		
$a_1^{(1)}$	1.000	/
$a_2^{(1)}$	0.787	[0.583;1.007]
$a_3^{(1)}$	1.656	[1.134;2.383]
$a_4^{(1)}$	0.000	/
$a_5^{(1)}$	0.140	[-0.057;0.330]
$a_6^{(1)}$	0.145	[-0.025;0.321]
Dimension 2		
$a_1^{(2)}$	0.000	/
$a_2^{(2)}$	0.124	[0.064;0.185]
$a_3^{(2)}$	-0.014	[-0.121;0.070]
$a_4^{(2)}$	1.000	/
$a_5^{(2)}$	1.124	[0.849;1.481]
$a_6^{(2)}$	0.882	[0.691;1.104]

Table 4.11: Estimation results of covariance matrix parameters : estimates and 95% credible intervals (CI) are presented.

	Est.	CI 95%
σ_1	1.756	[1.425; 2.154]
σ_2	0.038	[0.014; 0.065]
σ_3	3.959	[3.299; 4.697]
σ_4	0.079	[0.049; 0.112]
ρ_{12}	0.058	[-0.302; 0.462]
ρ_{13}	0.403	[0.240; 0.552]
ρ_{14}	0.229	[-0.079; 0.521]
ρ_{23}	-0.213	[-0.553; 0.106]
ρ_{24}	0.527	[0.100; 0.832]
ρ_{34}	-0.202	[-0.513; 0.185]

C7 Estimated distribution of response category probabilities

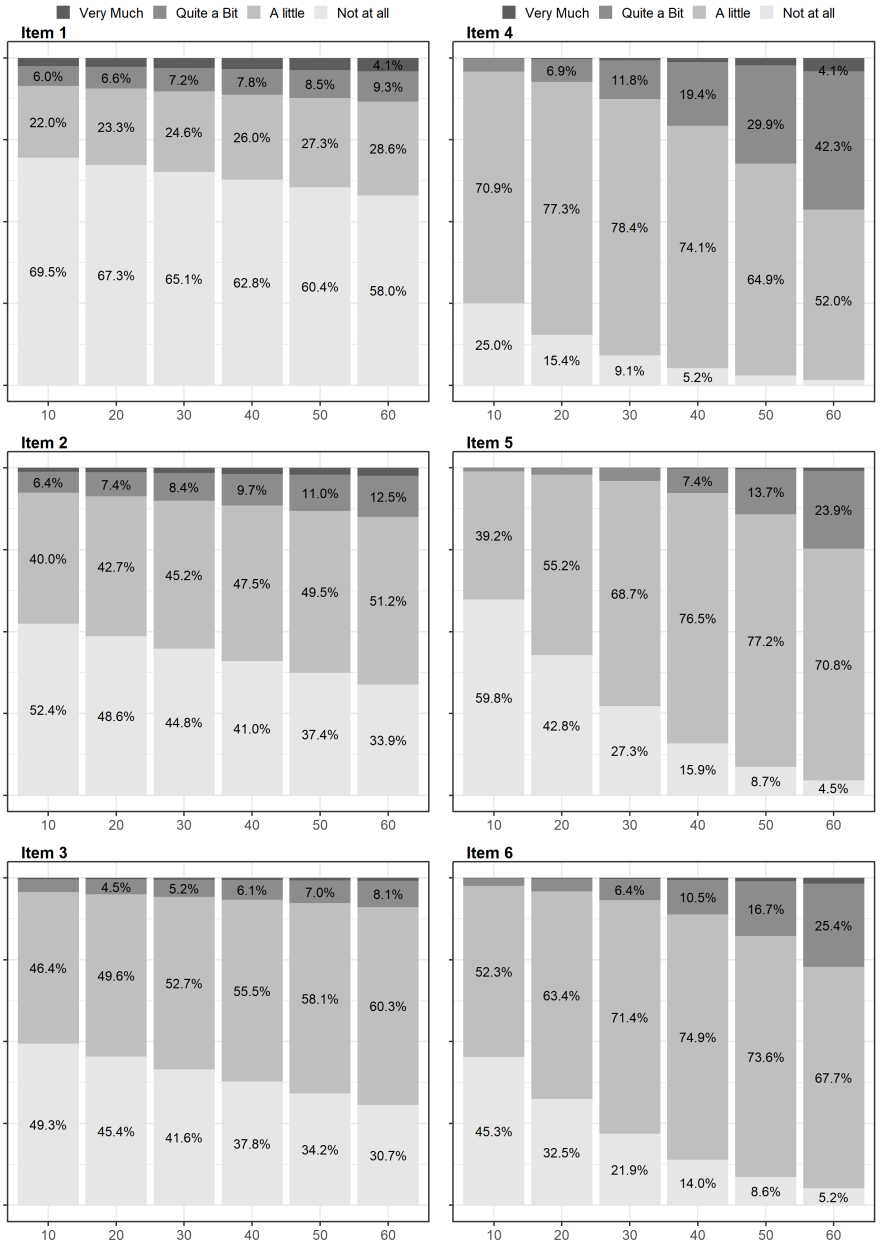

Figure 4.7: Estimated distribution of response category probabilities over time for all items.

Conclusion 5

This thesis develops methodological extensions for joint models of longitudinal outcomes and time-to-event data, motivated by challenges encountered in clinical trials. Particular attention is given to settings involving complex longitudinal structures, such as patient-reported outcomes, as well as to the presence of informative dropout in longitudinal data. The proposed contributions aim to improve the flexibility, interpretability, and practical applicability of joint modelling approaches in oncology research.

Main Contributions

The first contribution of this thesis focuses on allowing flexible effects of survival covariates in the survival submodel. In standard joint modelling approaches, the survival submodel is typically specified using a proportional hazards model with linear covariate effects. This assumption may be unrealistic and too restrictive in medical applications, where the effect of a covariate on the risk of event may be nonlinear.

We propose a joint modelling framework incorporating P-splines to model nonlinear covariate effects in the survival submodel, while allowing for non-Gaussian longitudinal outcomes through generalized linear mixed models.

The proposed approach, referred to as FlexibleJM, provides a more realistic and flexible representation of survival covariate effects while allowing accurate estimation of the parameters of the joint model. It is also important to note that the approach remains consistent when the true survival effects are in fact linear. This suggests that the method is worth considering for joint modelling of longitudinal and survival data when a nonlinear effect of a survival covariate is suspected. The approach is easily applicable and interpretable in practical applications, and its computation time is comparable to that of standard joint modelling methods.

The second contribution of this thesis was motivated by the growing importance of longitudinal health-related quality-of-life data in clinical trials. These outcomes are often analysed using simple sum score approaches, which have well-known limitations. This work aimed to contribute to a more appropriate

framework for analysing longitudinal HRQoL data through the use of item response theory. These longitudinal outcomes are often subject to high levels of missingness due to patient dropout, which may be informative and should be accounted for in the analysis.

To address these challenges, we introduced the JMIRT approach, a joint modelling framework that combines an item response theory model for multivariate longitudinal ordinal data with cause-specific hazard models for dropout. Unlike traditional joint models, which include the longitudinal process as a covariate in the survival submodel, the proposed approach shifts the emphasis to the longitudinal component by incorporating information from the dropout process directly into the latent trait model. The JMIRT framework therefore provides an efficient way to analyse longitudinal HRQoL data while accounting for different causes of dropout.

The application to HRQoL data from patients with progressive glioblastoma showed that this approach can capture clinically meaningful associations between dropout risks and patients' quality of life. These results highlight the importance of modelling dropout mechanisms jointly with longitudinal ordinal outcomes and may encourage more accurate reporting of dropout causes in future clinical trials.

However, these methodological advances in modelling multivariate ordinal longitudinal data within joint models involve a significant computational cost, especially when extending the framework to multidimensional latent traits. Indeed, HRQoL is often assessed across multiple dimensions, and analysing several latent dimensions simultaneously may be of particular interest. Fully joint Bayesian estimation of such models can quickly become impractical due to high computational burden.

The third contribution of this thesis addresses this issue by proposing a Slope-Corrected Two-Stage approach for multidimensional latent trait joint models. This approach builds on corrected two-stage methods by propagating information from the longitudinal model to the survival model through informative prior distributions on the random effects, while allowing longitudinal slope parameters to be re-estimated in the second stage in order to reduce potential bias.

Simulation studies demonstrate that the proposed slope-corrected two-stage approach substantially improves estimation compared with standard two-stage procedures, and its performance closely approaches that of fully joint Bayesian estimation, while requiring much lower computation time. Using this approach, multidimensional latent trait models can provide insights that

may be missed by univariate or score-based analyses, while remaining computationally efficient.

Overall, these three contributions address different methodological challenges encountered in the joint modelling framework for complex longitudinal and survival data in clinical trials. They also provide replicable solutions applicable to real-world clinical studies.

Methodological and Computational Considerations

Throughout this thesis, Bayesian estimation required the specification of prior distributions for all model parameters. Weakly informative priors were generally adopted, including normal priors for regression coefficients, Gamma-type priors for smoothing parameters in the P-spline components, and priors based on a variance–correlation decomposition of the random-effects covariance matrix.

Particular attention was given to the specification of covariance matrices. While inverse-Wishart priors have historically been widely used due to their conjugacy properties, allowing straightforward Gibbs updates, they may induce undesirable dependencies and unstable behaviour in complex or high-dimensional settings. For this reason, we instead rely on a decomposition into marginal variances and correlations, combined with weakly informative priors on standard deviations and LKJ priors (Lewandowski et al., 2009) on the correlation matrix, which generally leads to more stable and interpretable inference.

The impact of prior distributions depends on the amount of information available in the data. In settings with limited observations, priors may influence the estimates, whereas with sufficiently informative data, weakly informative priors allow the likelihood to dominate. Although no formal sensitivity analysis was conducted, the good estimation accuracy and coverage observed in the simulation studies suggest that the chosen prior priors are appropriate for the settings considered. Finally, these prior choices follow standard practice in the joint modelling and Bayesian smoothing literature.

A second aspect that deserves further discussion concerns computational challenges. Bayesian estimation of joint models can be computationally challenging, in particular with respect to the convergence and efficiency of Markov chain Monte Carlo algorithms. Throughout this thesis, several strategies were implemented to improve numerical stability and mixing of the chains. First, the evaluation of integrals involving the random effects in the survival sub-

model does not admit a closed-form expression under the proposed specifications and was therefore handled using numerical integration, relying on Gauss–Kronrod quadrature rules.

In addition, careful tuning of the Metropolis–Hastings algorithms was required. In particular, separately fitted longitudinal and survival models were used prior to the joint estimation to construct efficient covariance matrices for the proposal distributions, providing reasonable initial scaling and correlation structures. When necessary, adaptive strategies were also employed. For some parameters, empirical covariance estimates obtained from an initial phase of the Markov chains (e.g. based on the first 1000 iterations) were used to refine the proposal distributions and better capture posterior dependencies. In addition, proposal scales were adjusted during the algorithm using adaptive random-walk tuning schemes in order to target appropriate acceptance rates, which are known to improve the efficiency and convergence properties of Metropolis–Hastings algorithms.

Finally, in the context of the multidimensional longitudinal models considered in Chapter 4, the increasing complexity of the latent structure motivated the use of more advanced computational tools. In particular, the longitudinal submodel was implemented in `Stan`, which relies on Hamiltonian Monte Carlo methods. Unlike standard random-walk Metropolis–Hastings algorithms, these methods exploit gradient information from the posterior distribution to generate more efficient proposals, leading to improved convergence behaviour and more reliable estimation, especially for the variance–covariance matrix of the random effects, which is a key parameter in the methodology developed in this chapter.

Limitations

Despite the contributions of this thesis, some limitations should be noted.

First, the proposed flexible survival submodel estimates nonlinear covariate effects as population-averaged functions shared by all patients, which may mask clinically meaningful heterogeneity across subgroups. For example, the nonlinear effect of age on risk of death may differ between men and women, or between patients with low versus high comorbidity burden.

Second, in multidimensional graded response models, identifiability constraints must be imposed in order to avoid overparameterisation and to ensure proper model identification. These constraints are typically applied to item parameters and require fixing certain discrimination or threshold parameters for selected items within each dimension. While necessary, such constraints may

influence the interpretation of the resulting latent dimensions, especially in higher-dimensional settings where the choice of identification strategy becomes more delicate and often relies on prior substantive knowledge. Alternative identification strategies based on constraints imposed directly on the latent variables, such as fixing their means and covariances, are conceptually possible but challenging to implement in practice when multiple latent traits are involved, and may further complicate interpretation.

Finally, although the corrected two-stage strategies proposed in this thesis substantially reduce computation time compared with fully joint estimation, the longitudinal graded response model remains the most computationally demanding component of the analysis, as its estimation involves a large number of parameters. This limitation becomes more pronounced as the number of items or latent dimensions increases. In developing the corrected two-stage approaches, particular attention was paid to achieving a reasonable balance between computational efficiency and estimation accuracy. Indeed, the longitudinal model estimated in the first stage plays a central role, since key parameters obtained at this step, such as the variance of the random effects, are directly propagated to the second-stage survival model. It is therefore essential to obtain reliable estimates at this first stage. To the best of our knowledge, no existing estimation approach currently provides substantially faster inference for this class of longitudinal multidimensional graded response models while maintaining a comparable level of accuracy for all parameters of interest. For instance, the `mirt` package (Chalmers, 2012) allows slightly faster estimation of multidimensional graded response models, but it is currently not well suited to handle repeated longitudinal measurements jointly with mixed-effects structures for the latent traits.

Futures Perspectives

Beyond the contributions developed in this thesis, several methodological extensions within the joint modelling framework deserve further investigation. In particular, future work could focus on more complex ways of linking the longitudinal latent process to the survival outcome. In most joint models, the association is specified through the current value or the random effects of the longitudinal process. However, when the longitudinal outcome represents a latent health status, such as HRQoL, alternative formulations may be clinically more meaningful. For instance, the event risk may depend on the cumulative exposure to poor quality of life over time, on the rate of change of the latent trait, or on nonlinear relationships between the latent HRQoL process and the event risk. The latent trait itself may also be estimated flexibly to capture complex longitudinal patterns. Such extensions could provide a

more nuanced understanding of how longitudinal patient-reported outcomes relate to disease progression and survival.

Another important direction for future research concerns the assessment of measurement invariance in longitudinal item response theory models, in particular through the investigation of differential item functioning (DIF) and response shift (RS) (Teresi and Fleishman, 2007; Schwartz and Sprangers, 2010; Tay et al., 2015). IRT models rely on the assumption of measurement invariance, which states that individuals with the same level of the latent trait should have the same probability of endorsing an item, regardless of their characteristics or the time of assessment. In practice, this assumption may be violated in patient-reported outcome data, where patients may interpret questionnaire items differently depending on their profile or their experience of the disease.

Differential item functioning refers to situations in which individuals with the same latent trait level but different characteristics, such as clinical or demographic factors, show different response probabilities for specific items. Response shift is closely related but occurs within individuals, reflecting changes in the way questionnaire items are interpreted over time. Such changes may arise because patients reassess what different response levels mean to them, shift their personal values or priorities, or reconsider what the underlying concept represents for them as their experience of the disease evolves.

Extending the proposed joint modelling frameworks to explicitly account for DIF and RS, for instance by allowing item parameters to depend on covariates or time, would improve the validity of longitudinal HRQoL analyses. Beyond bias correction, modelling DIF and RS may also provide valuable information on how patients perceive and adapt to their health status over time.

Another promising perspective concerns the use of computerized adaptive testing (CAT) for the collection of longitudinal patient-reported outcomes (van der Linden and Glas, 2010). CAT is a computer-based assessment approach that relies on the IRT framework to adapt item administration to the individual, with the aim of maximising measurement precision while limiting patient burden. At each assessment occasion, items are selected sequentially based on the patient's previous responses and the current estimate of the latent trait, rather than administering a fixed set of items to all patients.

In the context of longitudinal PROs, the use of CAT may offer several advantages, including shorter questionnaires, improved patient compliance, and the possibility of more frequent assessments with limited additional burden. These features are particularly relevant in oncology studies, where patients

may experience fatigue or cognitive impairment. Importantly, although different patients may respond to different subsets of items, CAT-based data remain naturally compatible with item response theory models.

Recent developments illustrate the feasibility and relevance of this approach in clinical research. In particular, the EORTC has developed a CAT version of the QLQ-C30 questionnaire (Petersen et al., 2018). Extending the joint modelling frameworks considered in this thesis to accommodate CAT-based longitudinal data would represent a natural and promising direction for future research.

In conclusion, this thesis contributes to the development of joint modelling methodology by addressing several key challenges arising in the analysis of complex longitudinal and survival data from clinical trials. By successively extending the survival submodel, improving the analysis of multivariate ordinal longitudinal outcomes, and proposing computationally efficient estimation strategies, this work provides tools that are both statistically reliable and practically usable. The proposed perspectives open the way to further methodological developments for the analysis of patient-reported outcomes in current clinical research.

Bibliography

Alsefri, M., Sudell, M., Garcia-Finana, M., and Kolamunnage-Dona, R. (2020). Bayesian joint modelling of longitudinal and time to event data: a methodological review. *BMC Medical Research Methodology*, 20, 1-17.

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), 115.

Andersen, P. K., and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. The Annals of Statistics, 1100-1120.

Andrinopoulou, E. R., Rizopoulos, D., Takkenberg, J. J. M., and Lesaffre, E. (2014). Joint modeling of two longitudinal outcomes and competing risk data. *Statistics in Medicine*, 33, 3167–3178.

Andrinopoulou, E. R., Rizopoulos, D., Takkenberg, J. J. M., and Lesaffre, E. (2017). Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Statistical Methods in Medical Research*, 26, 1787–1801.

Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., Haes, J. C. J. M. de, Kaasa, S., Klee, M., Osoba, D., Razavi, D., Rofe, P. B., Schraub, S., Sneeuw, K. C. A., Sullivan, M., and Takeda, F. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute*, 85(5), 365–376.

Barbieri, A., Peyhardi, J., Conroy, T., Gourgou, S., Lavergne, C., and Mollevi, C. (2017). Item response models for the longitudinal analysis of health-related quality of life in cancer clinical trials. *BMC medical research methodology*, 17(1), 148.

Bottomley, A., Pe, M., Sloan, J., Basch, E., Bonnetain, F., Calvert, M., et al. (2016). Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. *The Lancet Oncology*, 17(11), e510–e514.

Brent, R. P. (1973). Some efficient algorithms for solving systems of nonlinear equations. *SIAM Journal on Numerical Analysis*, 10(2), 327–344.

Brown, E., and Ibrahim, J. G. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59, 221–228.

Brown, H., and Prescott, R. (2006). *Applied Mixed Models in Medicine* (2nd ed.). New York: John Wiley and Sons.

Brown, E. R., Ibrahim, J. G., and DeGruttola, V. (2005). A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61, 64–73.

Chahal, M., Thiessen, B., and Mariano, C. (2022). Treatment of older adult patients with glioblastoma: Moving towards the inclusion of a comprehensive geriatric assessment for guiding management. *Current Oncology*, 29, 360–376.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1–29.

Chi, Y., and Ibrahim, J. G. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62, 432–445.

ClinicalTrials.gov: National Library of Medicine (US). (2021). Identifier NCT01290939, Bevacizumab and Lomustine for Recurrent GBM, `https://clinicaltrials.gov/ct2/show/NCT01290939`.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187-202.

Crowther, M. J., and Lambert, P. C. (2013). Simulating biologically plausible complex survival data. *Statistics in Medicine*, 32(23), 4118–4134.

Crowther, M. J., Abrams, K. R., and Lambert, P. C. (2013). Joint modeling of longitudinal and survival data. *The Stata Journal: Promoting Communications on Statistics and Stata*, 13(1), 165–184.

Crowther, M. J., Andersson, T. M. L., Lambert, P. C., Abrams, K. R., and Humphreys, K. (2016). Joint modelling of longitudinal and survival data: incorporating delayed entry and an assessment of model misspecification. *Statistics in Medicine*, 35(7), 1193-1209.

Cuer, B., Conroy, T., Juzyna, B., Gourgou, S., Mollevi, C., and Touraine, C. (2022). Joint modelling with competing risks of dropout for longitudinal analysis of health-related quality of life in cancer clinical trials. *Quality of Life Research*, 1-12.

De Ayala, R. J. (2013). *The theory and practice of item response theory.* Guilford Publications.

Doms, H., Lambert, P., and Legrand, C. (2024). Flexible joint model for time-to-event and non-Gaussian longitudinal outcomes. *Statistical Methods in Medical Research*, 33(10), 1783–1799.

Doms, H., Lambert, P., and Legrand, C. (2025). Joint modeling of longitudinal HRQoL data accounting for the risk of competing dropouts. *arXiv preprint* arXiv:2503.03919. Manuscript submitted to *Journal of the Royal Statistical Society: Series C (Applied Statistics).*

Doms, H., Lambert, P., and Legrand, C. (submitted). Bias-Corrected Two-Stage Modeling for Joint Analysis of Multidimensional HRQoL and Survival. Manuscript submitted to *Pharmaceutical Statistics.*

ECOG-ACRIN Cancer Research Group. (2022). ECOG Performance Status. `https://ecog-acrin.org/resources/ecog-performance-status/`.

Eilers, P. H., and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2), 89–121.

Eilers, P., and Marx, B. (2021). *Practical Smoothing: The Joys of P-splines.* Cambridge University Press, Cambridge.

Elashoff, R. M., Li, G., and Lin, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*, 64, 762–771.

Elashoff, R., Li, G., and Li, N. (2016). *Joint Modeling of Longitudinal and Time-to-Event Data.* Chapman and Hall/CRC, Boca Raton.

Embretson, S. E. and Reise, S. P. (2013). *Item Response Theory for Psychologists.* Psychology Press.

Faucett, C., and Thomas, D. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine*, 15, 1663–1685.

Faucett, C. L., Schenker, N., and Elashoff, R. M. (1998). Analysis of censored survival data with intermittently observed time-dependent binary covariates. *Journal of the American Statistical Association*, 93, 427–437.

Fayers, P., Aaronson, N. K., Bjordal, K., Grønvold, M., Curran, D., and Bottomley, A. (2001). *EORTC QLQ-C30 scoring manual.* European Organisation for research and treatment of cancer.

Fayers, P. M., and Machin, D. (2013). *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes.* John wiley & sons.

Garcia-Hernandez, A., and Rizopoulos, D. (2018). %JM: a SAS macro to fit jointly generalized mixed models for longitudinal data and time-to-event responses. *Journal of Statistical Software*, 84, 1–29.

Geisser, S., and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74(365), 153-160.

Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457-472.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). Chapman and Hall/CRC, Boca Raton, FL.

Geweke, J. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments* (No. 148). Federal Reserve Bank of Minneapolis.

Gorlia, T., Stupp, R., Brandes, A. A., Rampling, R. R., Fumoleau, P., Dittrich, C., et al. (2012). New prognostic factors and calculators for outcome prediction in patients with recurrent glioblastoma: A pooled analysis of EORTC Brain Tumour Group phase I and II clinical trials. *European Journal of Cancer*, 48(8), 1176–1184.

Gorter, R., Fox, J. P., and Twisk, J. W. (2015). Why item response theory should be used for longitudinal questionnaire data analysis in medical research. *BMC Medical Research Methodology*, 15, 1-12.

Gould, A. L., et al. (2015). Joint modeling of survival and longitudinal non-survival data: current methods and issues. *Statistics in Medicine*, 34(14), 2181–2195.

Hays, R. D., Morales, L. S., and Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care*, 38(9 Suppl), II28–II42.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models.*. Chapman and Hall.

Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4), 465-480.

Holland, P.W. and Wainer, H. (2012). *Differential Item Functioning*. Routledge.

Hsieh, F.-C., Tseng, Y.-K., and Wang, J.-P. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, 62, 1037–1043.

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: Official Journal of the American Society of Clinical Oncology*, 28(16), 2796–2801.

Ibrahim J. G., Chen M-H, Sinha D. *Bayesian Survival Analysis*. Springer-Verlag; New York.

Jullion, A., and Lambert, P. (2007). Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. *Computational Statistics & Data Analysis*, 51(5), 2542–2558.

Klein, J., and Moeschberger, M. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer.

Köhler, M., Umlauf, N., Beyerlein, A., Winkler, C., Ziegler, A.-G., and Greven, S. (2017). Flexible Bayesian additive joint models with an application to type 1 diabetes research. *Biometrical Journal*, 59, 1144–1165.

Lambert, P., and Eilers, P. H. (2005). Bayesian proportional hazards model with time–varying regression coefficients: A penalized Poisson regression approach. *Statistics in Medicine*, 24(24), 3977-3989.

Lang, S., and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13(1), 183–212.

Legrand, C. (2021). *Advanced survival models*. Chapman and Hall/CRC.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9), 1989-2001.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. 2nd edition, Wiley, New York.

Luo, S. (2014). A Bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time. *Statistics in Medicine*, 33(4), 580-594.

Marx, B. D., and Eilers, P. H. C. (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics & Data Analysis*, 28, 193–209.

Mauff, K., Steyerberg, E., Kardys, I., Boersma, E., and Rizopoulos, D. (2020). Joint models with multiple longitudinal outcomes and a time-to-event outcome: a corrected two-stage approach. *Statistics and Computing*, 30(4), 999-1014.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13(2), 127–143.

Moore, D. F. (2016). *Applied Survival Analysis Using R*. Springer, Switzerland.

National Cancer Institute, NIH, DHHS, Bethesda, MD, April 2025, `https://progressreport.cancer.gov`.

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(1), 223-240.

Petersen, M. A., Aaronson, N. K., Arraras, J. I., Chie, W.-C., Conroy, T., Costantini, A., Dirven, L., Fayers, P., Gamper, E.-M., Giesinger, J. M., Habets, E. J. J., Hammerlid, E., Helbostad, J., Hjermstad, M. J., Holzner, B., Johnson, C., Kemmler, G., King, M. T., Kaasa, S., Loge, J. H., and Groenvold, M. (2018). The EORTC CAT Core—The computer adaptive version of the EORTC QLQ-C30 questionnaire. *European Journal of Cancer*, 100, 8–16.

Proust-Lima, C., Joly, P., Dartigues, J. F., and Jacqmin-Gadda, H. (2009). Joint modelling of multivariate longitudinal outcomes and a time-to-event: a nonlinear latent class approach. *Computational Statistics & Data Analysis*, 53(4), 1142-1154.

Proust-Lima, C., Philipps, V., and Liquet, B. (2017). Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. *Journal of Statistical Software*, 78, 1-56.

Putter, H., Fiocco, M., and Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi–state models. *Statistics in Medicine*, 26(11), 2389-2430.

Rizopoulos, D. (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35, 1-33.

Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press.

Rizopoulos, D. (2012). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics & Data Analysis*, 56(2), 491–501.

Rizopoulos, D. (2016). The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *Journal of Statistical Software*, 72, 1-46.

Rizopoulos, D. (2023). *JMbayes2: Joint Models for Longitudinal and Time-to-Event Data under the Bayesian Approach*. R package version 0.6. Available at: `https://CRAN.R-project.org/package=JMbayes2`.

Rizopoulos, D., and Ghosh, P. (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine*, 30, 1366–1380.

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika*, 95(1), 63–74.

Rondeau, V., Mazroui, Y., and Gonzalez, J. R. (2007). Joint frailty models for recurrent events and death using maximum penalized likelihood estimation: application to cancer events. *Biostatistics*, 8(4), 708–721.

Rondeau, V., Marzroui, Y., and Gonzalez, J. R. (2012). frailtypack: an R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software*, 47(1), 1-28.

Rue, H., and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. CRC Press, Boca Raton.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, 34(S1), 1-97.

Saulnier, T., Philipps, V., Meissner, W. G., Rascol, O., Pavy-Le Traon, A., Foubert-Samier, A., and Proust-Lima, C. (2022). Joint models for the longitudinal analysis of measurement scales in the presence of informative dropout. *Methods*, 203, 142-151.

Schwartz, C. E., and Sprangers, M. A. G. (2010). Guidelines for improving the stringency of response shift research. *Quality of Life Research*, 19, 455–464.

Self, S., and Pawitan, Y. (1992). Modeling a marker of disease progression and onset of disease. In *AIDS epidemiology: methodological issues* (pp. 231-255). Boston, MA: Birkhäuser Boston.

Singer, M., and Webb, A. R. (2005). *Oxford Handbook of Critical Care*. Oxford University Press.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583-639.

Sprangers, M. A., and Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: a theoretical model. *Social Science & Medicine*, 48(11), 1507-1515.

Taal, W., et al. (2014). Combination of bevacizumab and lomustine in patients with first recurrence of glioblastoma. *Journal of Clinical Oncology*, 32(12), 1239–1247.

Tan, A. C., Ashley, D. M., López, G. Y., Malinzak, M., Friedman, H. S., and Khasraw, M. (2020). Management of glioblastoma: State of the art and future directions. *CA: A Cancer Journal for Clinicians*, 70(4), 299–312.

Tay, L., Meade, A. W., and Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods*, 18(1), 3–46.

Tennant, A. and Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? *Rheumatology*, 46(8), 1348–1352.

Teresi, J. A., and Fleishman, J. A. (2007). Differential item functioning and health assessment. *Quality of Life Research*, 16(Suppl 1), 33–42.

Therneau, T., and Grambsch, P. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York.

Therneau, T. (2023). *A Package for Survival Analysis in R*. R package version 3.8.3. Available at: https://CRAN.R-project.org/package=survival.

Therneau, T. (2026). *Spline terms in a Cox model*. R package `survival` vignette. Available at: https://cran.r-project.org/web/packages/survival/vignettes/splines.pdf.

Thiébaut, R., Jacqmin-Gadda, H., Babiker, A., Commenges, D., and Cascade Collaboration. (2005). Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of CD4+ cell count and HIV RNA viral load in response to treatment of HIV infection. *Statistics in Medicine*, 24(1), 65-82.

Touraine, C., Cuer, B., Conroy, T., Juzyna, B., Gourgou, S., and Mollevi, C. (2023). When a joint model should be preferred over a linear mixed model for analysis of longitudinal health-related quality of life data in cancer clinical trials. *BMC Medical Research Methodology*, 23(1), 36.

Tsiatis, A. A., Degruttola, V., and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*, 90(429), 27-37.

Tsiatis, A. A., and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14(3), 809–834.

Umlauf, N., Klein, N., and Zeileis, A. (2018). BAMLSS: Bayesian additive models for location, scale, and shape (and beyond). *Journal of Computational and Graphical Statistics*, 27(3), 612-627

van der Linden, W. J., and Glas, C. A. W. (2010). *Elements of Adaptive Testing*. New York: Springer.

Van der Linden, W. J. (2016). *Handbook of item response theory*. New York: CRC press.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.

Verhagen, J., and Fox, J. P. (2013). Longitudinal measurement in health–related surveys. A Bayesian joint growth model for multivariate ordinal responses. *Statistics in Medicine*, 32(17), 2988-3005.

Villani, C. (2009). *Optimal transport: old and new*. Berlin: springer.

Viviani, S., Alfò, M., and Rizopoulos, D. (2014). Generalized linear mixed joint model for longitudinal and survival outcomes. *Statistics and Computing*, 24(3), 417–427.

Wang, J., and Luo, S. (2017). Multidimensional latent trait linear mixed model: an application in clinical studies with multivariate longitudinal outcomes. *Statistics in Medicine*, 36(20), 3244-3256.

Wang, J., and Luo, S. (2019). Joint modeling of multiple repeated measures and survival data using multidimensional latent trait linear mixed model. *Statistical Methods in Medical Research*, 28(10-11), 3392-3403.

Wang, Y., and Taylor, J. (2001). Jointly modeling longitudinal and event time data with application to AIDS. *Journal of the American Statistical Association*, 96(455), 895–905.

Wang, J., Luo, S., and Li, L. (2017). Dynamic prediction for multiple repeated measures and event time data: An application to Parkinson's disease. *The Annals of Applied Statistics*, 11(3), 1787–1809.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.

Wick, W., Gorlia, T., Bendszus, M., Taphoorn, M., Sahm, F., Harting, I., et al. (2017). Lomustine and bevacizumab in progressive glioblastoma. *New England Journal of Medicine*, 377(20), 1954–1963.

Williamson, P., Kolamunnage-Dona, R., Philipson, P., and Marson, A. (2008). Joint modelling of longitudinal and competing risks data. *Statistics in Medicine*, 27, 6426–6438.

Wood, S. N. (2013). On P-values for smooth components of an extended generalized additive model. *Biometrika*, 100(1), 221–228.

Wood, S. (2017). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC.

World Health Organization. (2024, February 1). *Global cancer burden growing, amidst mounting need for services*. News release. `https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services`.

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(1).

Wulfsohn, M., and Tsiatis, A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1), 330–339.

Ye, W., Lin, X., and Taylor, J. M. (2008). Semiparametric modeling of longitudinal measurements and time-to-event data–a two-stage regression calibration approach. *Biometrics*, 64(4), 1238-1246.