Multilevel Bayesian Joint Model in Hierarchically Structured Data

A dissertation submitted to the

Graduate School

of the University of Cincinnati

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Department of Mathematical Sciences

of the College of Arts and Sciences

by

Chen(Grace) Zhou

M.S. University of Texas at Dallas

July 2022

Committee:

Seongho Song, Ph.D., Chair

Won Chang, Ph.D.

Hang Joon Kim, Ph.D.

Rhonda D. Szczesniak, Ph.D.

Xia Wang, Ph.D.

Abstract

Joint modeling has been a useful strategy for incorporating latent associations between different types of outcomes simultaneously in the last two decades. This dissertation contributes to the development of a multilevel Bayesian joint model, which is motivated by a longitudinal lung disease study. In Chapter 1, the background of Bayesian methodology for the joint modeling is introduced. Chapters 2 and 3 describe two novel joint models with applications to the multi-center data for cystic fibrosis disease. First, in Chapter 2, a multilevel Bayesian joint model of longitudinal continuous and binary outcomes is proposed. Second, in Chapter 3, a multilevel Bayesian joint model of longitudinal and recurrent outcomes is postulated. Lastly, in Chapter 4, some key takeaways, limitations and future work are discussed.

To the loving memory of my grandfather

Acknowledgments

That my God would grant me, according to the riches of His glory, to be strengthened with power through His Spirit into my inner man (Eph. 3:16), so that my Ph.D. program at the University of Cincinnati (UC) can finally be accomplished as it should be. Specifically, Jesus Christ has settled a cloud of people surrounding me, providing unlimited helps and encouragements.

I would like to express my deepest appreciation to my academic advisor, Professor Seongho Song, for guiding and shaping my research interest throughout my Ph.D. study. My sincere appreciation also goes to my co-advisor (also mentor), Professor Rhonda D. Szczesniak, whose insights and experiences greatly steer me through considerable researches at Cincinnati Children's Hospital Medical Center (CCHMC).

I am also grateful to the committee members, Drs. Won Chang, Hang Joon Kim and Xia Wang for their valuable and helpful comments. In addition, I would like to thank Dr. Yizao Wang, who is an excellent professor full of supports to students.

A word of special thanks goes to all faculty & staff members in the Department of Mathematical Sciences at UC. I deeply appreciate the generous scholarship sponsored by my department (Division of Statistics and Data Science) and well-established learning system along with many useful resources, such as study room, computer lab, online library resources, academic writing center, etc. In addition, the public library at Clifton (Cincinnati, OH) and the public library at Richardson (Richardson, TX) and the study lounge at University of

Texas at Dallas (UTD) all provide me a great place to concentrate and be productive. Meanwhile, I would like to extend my thanks to the internship at Procter & Gamble, my supervisor Bambi Rosenthal, who provided me the opportunity to build a solid skill in R Shiny apps.

Most importantly, I give my biggest thanks to my families, specially to my dearest parents and grandparents, for their unconditional loves and unreserved cares throughout my life span. What's more, two furry friends, Lillian and Mooyah, deserve lots of treats for their 24/7 accompany as a barrel of laughs.

Last but not the least, I want to express my sincere gratitude to my spiritual parents Brother David Chen and Sister Rainbow Gan, for their unceasing prayers and shepherding. Also sister May Hsu, for her hospitality in kindness and love. I am grateful to have many friends, such as Esther Lin, Eunsun Yook, Anushka Palipana, an anonymous friend, etc, who always stand by me.

The research of this dissertation was supported by grant R01 HL141286 from the National Institutes of Health. The paper presentation for Chapter 3 at 2022 Joint Statistical Meetings (JSM) is sponsored by Graduate Student Government (GSG) Research Fellowship Awards.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter introduces the background of joint modeling, estimation approaches and Bayesian hierarchical model. To demonstrate an efficient sampling method, we illustrate an example of hierarchical model. Given the feature of the motivating data, we explain the levels of hierarchy, primary outcomes and motivations.

1.1 Background for Joint Model

1.1.1 Classic Joint Model

In many clinical or epidemiological research studies, the longitudinal data may be censored by a time-to-event outcome, such as death or dropout. In order to explore how changes in the biomarker are associated with the occurrence of the terminal event, joint modeling of longitudinal and survival data has became to be prevailing (Tsiatis and Davidian (2004),

Henderson et al. (2000), Rizopoulos (2012a), Ibrahim et al. (2010), Wulfsohn and Tsiatis (1997), Asar et al. (2015)). It was summarized in three main features for its popularity (Hickey et al. (2016)):

1. Improving inference for a repeated measurement outcome subject to an informative dropout mechanism;

2. Improving inference for a time-to-event outcome, whilst taking account of an intermittently error-prone measured endogenous time-dependent variable;

3. Studying the relationship between the two correlated processes

Generally, a joint model with shared parameter refers to the simultaneous estimation of a longitudinal outcome characterized by a linear mixed effects (LME) submodel and a terminal event subject to a survival submodel. Such two processes are linked using shared individual-specific parameters, which can be parameterised in a number of ways, such as random effects (see Chapter 2), current value (see Chapter 3), time-dependent slope (see Chapter 3), cumulative effects, lagged time, etc.

1.1.2 Extended Joint Model

The joint model has focused on modeling a single longitudinal outcome and a single time-to-event outcome; thereby known as univariate joint modeling. Nonetheless, a vast number of extensions have been proposed to increase the flexibility for complicated studies, such as

latent class joint model (Proust-Lima et al. (2014)), competing risks (Andrinopoulou et al. (2017)), recurrent events (Król et al. (2016), Hickey et al. (2018a), Ren et al. (2021)), multiple longitudinal outcomes (Musoro et al. (2015)), accelerated failure time (Luo (2015)), longitudinal binary outcomes (Horrocks and van Den Heuvel (2009)). Among those extensions, we are interested in joint model for binary outcomes (see Chapter 2) and joint model for time-to-recurrent events (see Chapter 3) given the feature of our motivating data. Moreover, joint modeling has been demonstrated to improve the prediction (Rizopoulos et al. (2014)). Our prognostic rationale is primarily based on inherently Bayesian-frequentist approach in Chapter 2 and predictive posterior distribution incorporated with Monte Carlo method in see Chapter 3.

1.1.3 Key Assumption

The key assumption for joint modeling is the conditional independence, which means random effects explain all the interdependence. So that given random effects,

- Submodels are mutually independent
- Repeated measurements/events in each submodel are independent of each other

1.1.4 Parameter Estimation

Frequentist and Bayesian estimations are utilized as two widespread approaches to manage technical and computational challenges from aforementioned joint models. The Bayesian

statistics is older than frequentist statistics, but it has been neglected over the years due to computer technologies and discovery of new mathematical methods (Hackenberger (2019)).

A class of algorithms from notion of Markov Chain Monte Carlo (MCMC) (Metropolis et al. (1953), Hastings (1970), Geman and Geman (1984), Tanner and Wong (1987), Gelfand and Smith (1990), Geyer (1992), Tierney (1994)) initiated the era of modern Bayesian approach, which offers insight into estimating from posterior probability density functions that are not analytically tractable. For a comprehensive treatment of MCMC techniques, we defer readers to the handbook of MCMC (Brooks et al. (2011)). Furthermore, Bayesian statistics incorporates the ease of hierarchical models, such as employed in Luo and Wang (2014) and Brilleman et al. (2019). Consequently, main methodologies addressed in this dissertation are in Bayesian perspective, nonetheless, we adapt to joint Bayesian-frequentist approach for dynamic individual prediction as appropriate in Chapter 2.

1.1.5 Computation Method

A number of packages from mainstream statistical softwares, including R (R Core Team (2020)), SAS (SAS Institute, Cary, NC), Stata (Stata-Corp LP, College Station, TX) and WinBUGS (MRC Biostatistics Unit, Cambridge, UK), have allowed researchers to well exploit joint modeling (Hickey et al. (2016)). In this dissertation, we are primarily dependent on R software, thus summarize existing R implementation in the Table 1.1.

The recently released JMBayes2 (Rizopoulos et al. (2022)) is an extension from JMBayes

(Rizopoulos (2016)), which is powerful in fitting extended joint models via MCMC implemented in C++. Another Bayesian user-friendly R package named rstanarm utilizes Stan (Goodrich et al. (2020)) for the back-end estimation. The convenient stan_jm function allows for univariate or multivariate joint model with avoidance of writing own Stan programs (Brilleman et al. (2018)). Despite the robustness of existing R packages, we implement our own R codes along with Stan programs to meet the particular need for our motivating data and example codes can be downloaded from my Github (see links in Appendix B.6 & Appendix C.7).

Table 1.1: Implementations from R packages

R package	**Method**	**Bayesian**	**Reference**
joineR	Expectation Maximization algorithm	×	Philipson et al. (2018)
joineRML	Monte Carlo Expectation Maximization algorithm	×	Hickey et al. (2018b)
JM	Maximum Likelihood Estimation	×	Rizopoulos (2010)
lcmm	Maximum Likelihood Estimation	×	Proust-Lima et al. (2022)
frailtypack	Maximum Penalized Likelihood Estimation	×	Rondeau et al. (2012) Rondeau et al. (2019)
JMBayes	Monte Carlo Markov Chain (MCMC)	✓	Rizopoulos (2016)
JMBayes2	MCMC	✓	Rizopoulos et al. (2022)
rstanarm *	Hamiltonian Monte Carlo	✓	Brilleman et al. (2018)

* function stan_jm()

1.2 Background for Hierarchical Model

Throughout this dissertation, we take the key idea of the hierarchical (multilevel) model, allowing for observable outcomes being modeled conditionally on certain parameters, which themselves are given a probabilistic specification in terms of further parameters, known

as hyperparameters. Hierarchical concept helps in understanding multiparameter problems and also plays an important role in developing computational strategies. More importantly, simple nonhierarchical models are usually inappropriate for hierarchical data. (Gelman et al. (2013a))

1.2.1 Bayesian Inference

Let $\boldsymbol{\theta}$ denote unknown parameters with $\boldsymbol{\phi}$ as the corresponding hyperparameters and $\boldsymbol{y}$ represent observed data. The key 'hierarchical' part is that $\boldsymbol{\phi}$ is unknown and thus has its own prior distribution, $p(\boldsymbol{\phi})$. The joint posterior distribution is given by

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\phi})p(\boldsymbol{\theta}, \boldsymbol{\phi}) \\ &= \underbrace{p(\boldsymbol{y}|\boldsymbol{\theta})}_{\text{likelihood}} \quad \underbrace{p(\boldsymbol{\theta}|\boldsymbol{\phi})}_{\text{1st-stage prior}} \quad \underbrace{p(\boldsymbol{\phi})}_{\text{2nd-stage prior}} \end{aligned} \tag{1.1}$$

with the latter simplification holding because data distribution depends only on $\boldsymbol{\theta}$. The prior distribution for $\boldsymbol{\phi}$ relies on its extent of uncertainty. Either diffuse prior or (weakly) informative prior can be utilized, however, we should ensure the resulting posterior distribution is proper when using an improper prior density. More details of prior choice recommendations defer to Gelman's post(Gelman (2020)).

1.2.2 Hamiltonian Monte Carlo

Although the primary implementation is on the ubiquitous stochastic MCMC methods, in particular Metropolis-Hastings (Hastings (1970)) and Gibbs sampling (Gelfand and Smith (1990)), Betancourt and Girolami (Betancourt and Girolami (2013)) introduced a sophisticated but novel MCMC techniques, which is well known as Hamiltonian Monte Carlo (Neal (2011)). It utilizes techniques from differential geometry to generate transitions spanning the full marginal variance, eliminating the random walk behavior endemic to Random Walk Metropolis and the Gibbs samplers, therefore, provides the efficient exploration for the complex hierarchical models. Detailed algorithms are not included here due to the distraction from our main purpose. Readers can refer to the Section 3 of their paper for theoretical interests.

1.2.3 Probabilistic Programming Language

The inference engine Stan (not an acronym; Stan Development Team (2011-2019)) releases computational constraints of Euclidean HMC. Users can build, test and run hierarchical models through this powerful probabilistic programming language for specifying the target function. Particularly, Stan adapts to No-U-Turn Sampler to preserves detailed balance by integrating not just forward in time but also backwards (Hoffman and Gelman (2011)).

In this section, we illustrate an example of hierarchical model in Stan through the combinations of two interfaces and two methods of parameterizations (centered and non-

centered). Specifically, the interfaces to Stan are described as a new lightweight interface named CmdStanR(Gabry and Cešnovar (2022)) and the conventional interface named RStan (Stan Development Team (2020)). Note that in order to execute CmdStanR, it is necessary to install CmdStan first. Then we consider the Eight Schools example (Rubin (1981)) in terms of the centered parameterization (Betancourt (2017)),

$$y_n \sim N(\theta_n, \sigma_n)$$

$$\theta_n \sim N(\mu, \tau)$$

$$\mu \sim N(0, 5)$$

$$\tau \sim \text{Half-Cauchy}(0, 5)$$

where $n \in \{1, \ldots, 8\}$ and $\{y_n, \sigma_n\}$ are given data. Secondly, with respect to the non-centered parameterization, we introduce a latent Gaussian variable from which we can recover the group-level parameters with a scaling and a translation,

$$y_n \sim N(\theta_n, \sigma_n)$$

$$\tilde{\theta} \sim N(0, 1)$$

$$\mu \sim N(0, 5)$$

$$\tau \sim \text{Half-Cauchy}(0, 5)$$

$$\theta_n = \mu + \tau \cdot \tilde{\theta}$$

In practice, we have the same setup for the four scenarios, for instance, seed=2022, adapt_delta=0.8, max_treedepth=10, chains=2, warmup=1000, iters=2000. The Stan programs and R code are included in Appendix A.1. From Figure 1.1, we observe that resulting mean of $log(\tau)$ is strongly biased away from the true value for centered parameterization. Nonetheless, the pathology of bias and divergence is well solved by non-centered method and a thorough discussion of the non-centered intuition can be found in Betancourt and Girolami (2015). It is worth noting that both interfaces are supposed to bear the same results as long as all inputs are the same. In this motivating example, we allow initial values to be unassigned so that each trajectory is distinguishable and we aim to compare elapsed time and system compatibility rather than doubting the estimates.

Figure 1.2 shows the chains from centered parameterization (left panel) are sticking as they approaches negative values of $log(\tau)$ (or small values of τ), which is indicative of the divergences. On the contrary, non-centered parameterization (right panel) stretches $log(\tau)$ further towards negative values and shows the capability to explore the neck of the funnel.

To the end, we apply the superior non-centered trick to our proposed joint models in Chapter 2 & 3. Besides, we confirm less elapsed time by interface CmdStanR, although such advantage is not obvious in this example. However, when we have complicated model setups and larger sample size, the efficiency will be easily observed (see Appendix B.5). We also note that RStan is more compatible to windows system, while CmdStanR is more favorable to macOS system. For applications, RStan is mainly utilized in Chapter 2 and CmdStanR is employed in Chapter 3. To access the accuracy of posterior mean estimation,

equal-tailed credible intervals are employed. Notwithstanding 95% credible interval (CI) from RStan is common, 90% posterior uncertainty intervals from CmdStanr are also recommended due to more computational stability and relation to Type-S errors (Gelman and Carlin (2014)). Both intervals are incorporated in our dissertation, whilst taking account of a simulation study with S replicates, we compute its corresponding 95% confidence interval for an averaged posterior mean estimate, that is $\bar{\theta} \pm z_{0.025} * \frac{\sqrt{\sum_{s=1}^{S}(\hat{\theta}_s - \bar{\theta})^2/(S-1)}}{\sqrt{S}}$ with $\bar{\theta} = \frac{\sum_{s=1}^{S}\hat{\theta}_s}{S}$.

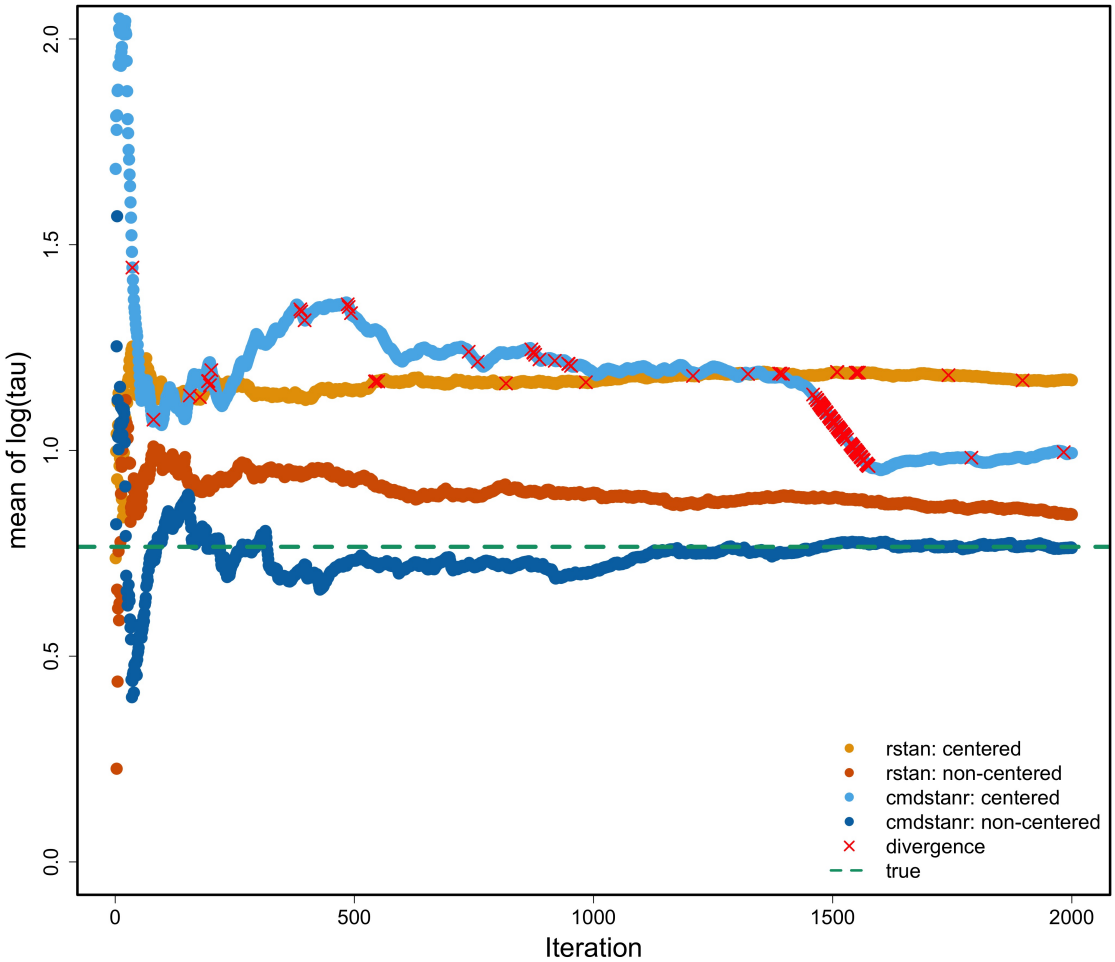

Figure 1.1: Convergence of mean $log(\tau)$ towards the true expected value across four scenarios

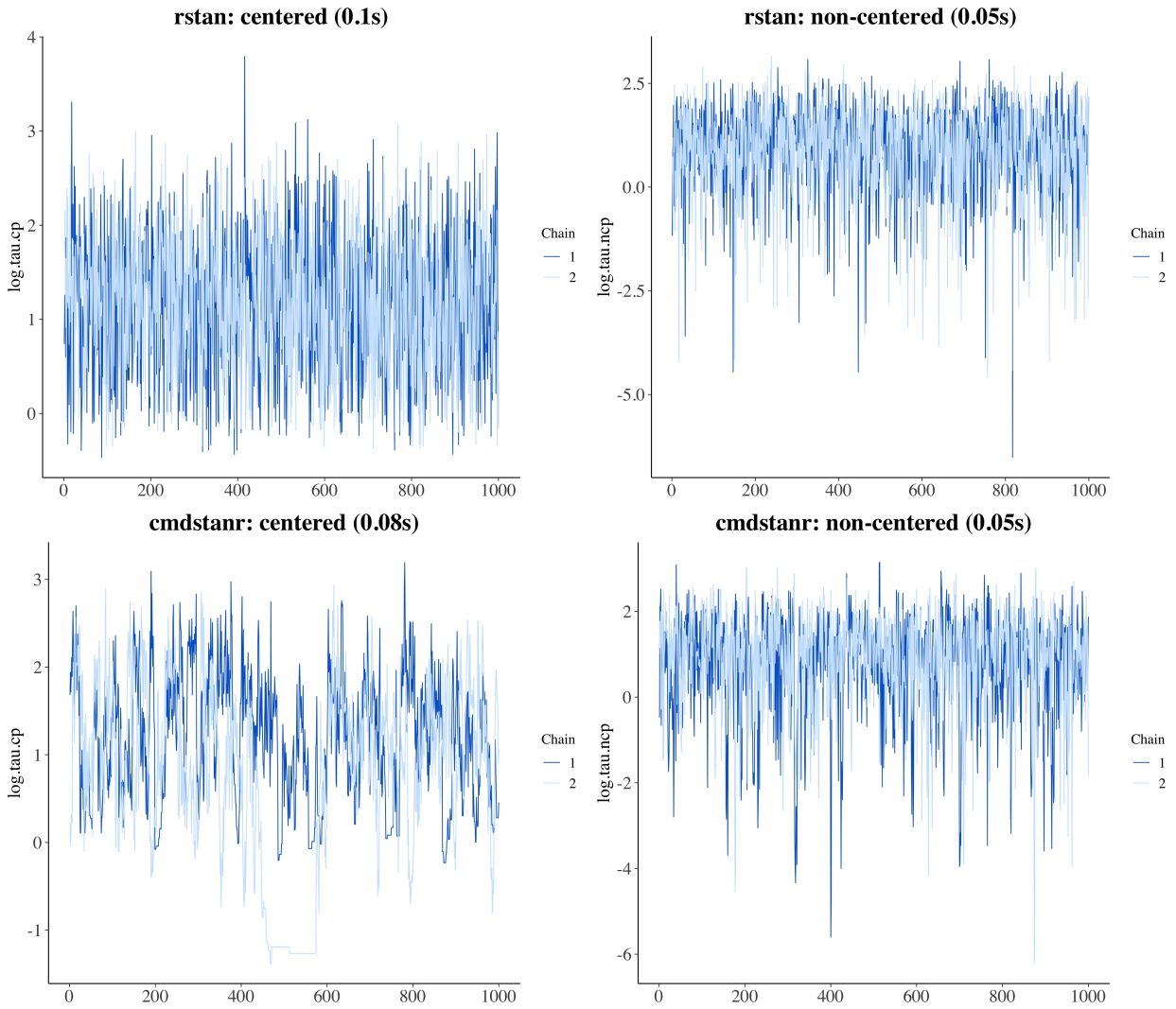

Figure 1.2: Trace plot of $log(\tau)$ across four scenarios with elapsed time in parentheses

1.3 Main Motivation

The main motivation of this dissertation arises from the personalized prediction of pulmonary exacerbation (PEx) risk in a epidemiological study of United States Cystic Fibrosis Foundation Patient Registry (US CFFPR). In order to monitor progressive changes of lung function

in people living with cystic fibrosis (CF), a variety of measures and outcomes are collected at each patient's clinical visit. Typically, a primary longitudinal marker named percent predicted forced expiratory volume in 1 second (hereafter, ppFEV1), is commonly used to describe the severity of lung disease. This continuous outcome is obtained from Global Lung Initiative Equations based on observed FEV1 in liters (Quanjer et al. (2012)). Whilst the binary event PEx is defined by a recorded antibiotics treatment during hospitalization. One inconceivable feature for our registry data is of three-level hierarchical structure, which is displayed in Figure 1.3; longitudinal measurements of the biomarker ppFEV1 and PEx events are observed at time points (1st level of the hierarchy), which are clustered within patients (2nd level of the hierarchy), who come from local CF centers (3rd level of the hierarchy). We also call such data structure as the multi-center data. For this cause, we are motivated to propose a multilevel Bayesian joint model to study the underlying associations between the two processes of ppFEV1 and PEx. Specifically, we regard PEx as a binary longitudinal outcome in Chapter 2 and a time-to-recurrent event in Chapter 3.

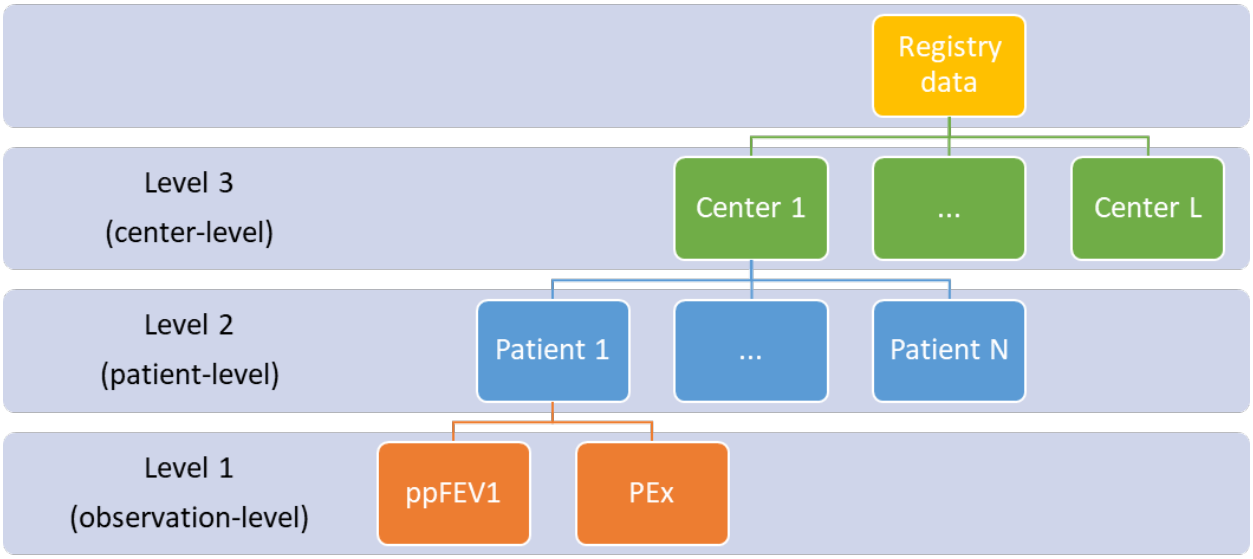

Figure 1.3: Example of the hierarchical structure of CF registry data

Chapter 2

Multilevel Bayesian joint model of longitudinal continuous and binary outcomes

In this chapter, we regard PEx as the longitudinal binary outcome by proposing a multilevel Bayesian joint model (JM) that encompasses the Linear Mixed Effect (LME) model with a Gaussian process and the Generalized Linear Mixed Model (GLMM) with a flexible link function. This chapter is organized as follows. In Section 2.1, we address the motivation. In Section 2.2, we introduce the methodology, including model framework, Bayesian inference, predictive metrics and model comparison criterion. In Section 2.3, we illustrate two simulation studies, of which one is for comparison between the flexible link and fixed links,

another one is about exploring four model structures under the flexible link. In Section 2.4, we apply the proposed joint model to a motivating example. Lastly, we conclude our study with remarks and discussions in Section 2.5.

2.1 Motivation

Although joint modeling has been a useful strategy in modern longitudinal analysis, applications to hierarchical longitudinal studies have been less frequent, particularly with respect to a binary process, which is commonly specified by a Generalized Linear Mixed Model (GLMM). Moreover, an added complexity occurs when the Gaussian process for irregular visits and flexible link function are involved to facilitate estimation and prediction. To the best of our knowledge, these two key challenges have not been considered together in the existing joint modeling approaches.

For this cause, we propose a joint model that unites an LME submodel for the evolution of ppFEV1 and a GLMM for the occurrence of PEx. Both center- and individual-level random intercepts are the basis of latent associations between these two submodels. In addition, we employ a stationary exponential correlation function for ppFEV1 to allow for longstanding correlation and intrinsic nonlinearity (Szczesniak et al. (2017)) and apply flexible link functions (Jiang et al. (2013)) for PEx occurrence, which comprise both symmetric and asymmetric link functions to capture the underlying skewness of responses. To summarize, our motivation is threefold: i) compare the flexible link with three common links; ii)

quantify biases caused by non-hierarchical joint model for a multicenter cohort; iii) evaluate prognostic utility of the optimal joint model.

2.2 Model Framework

2.2.1 The Extended Joint Model

Our proposed shared parameter joint model is expressed as

$$\begin{cases} Y_{lij} = \boldsymbol{X}_{li}(t_{lij})\boldsymbol{\alpha} + b_l + U_{li} + W_{li}(t_{lij}) + \epsilon_{lij}, \\ Pr(R_{lij} = 1) = g^{-1}\big(\boldsymbol{V}_{li}(t_{lij})\boldsymbol{\beta} + \rho_1 b_l + \rho_2 U_{li}\big) \end{cases} \tag{2.1}$$

where Y_{lij} denotes longitudinal continuous profiles and R_{lij} represents longitudinal binary response observed at time point t_{lij} for subject i from center l, with $l = 1, \ldots, L; i = 1, \ldots, n_l; j = 1, \ldots, n_{li}$. $\boldsymbol{X}_{li}(t_{lij})$ and $\boldsymbol{V}_{li}(t_{lij})$ denote row vectors of explanatory variables and $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are corresponding unknown coefficients. Between-center and between-patient heterogeneity are incorporated by random intercept terms b_l and U_{li}, allowing for the assumption to be independent and identically distributed (i.i.d.) from $N(0, \sigma_b^2)$ and $N(0, \sigma_u^2)$, respectively. A random-intercept model rather than the random intercept-and-slope model (Laird and Ware (1982)) is adapted, because the latter is too rigid to capture the pattern of variability in ppFEV1 over longer periods of time (Taylor-Robinson et al. (2012)). The stochastic process $\boldsymbol{W}_{li}(t)$ describes how individual lung function varies over time and is assumed to be independent copies of a zero-mean, continuous-time stationary Gaussian process with the

covariance function defined as

$$Cov(W_{li}(t), W_{li}(s)) = \tau^2 exp\big(- |t - s| \cdot \rho\big) \tag{2.2}$$

where t and s are two arbitrary time points, ρ represents the inverse of a range parameter and τ is the scale parameter for exponential correlation. A concept for nugget factor $1 -$ nugget $= \frac{\tau^2}{\sigma^2+\tau^2}$ incorporates both scale parameters τ and σ, with measurement error $\boldsymbol{\epsilon}$ following $N(0, \sigma^2)$. Furthermore, we assume that $\boldsymbol{b}$, $\boldsymbol{U}$ and $\boldsymbol{\epsilon}$ are mutually independent.

Let R_{lij} be a Bernoulli random variable with probability $Pr(R_{lij} = 1)$. $g(\cdot)$ denotes some known link function, which is composed of fixed effects $\boldsymbol{V}_{li}(t_{lij})\boldsymbol{\beta}$ and shared random components b_l, U_{li}. Parameters ρ_1 and ρ_2 measure the strength of the association between the two submodels.

2.2.2 Symmetric Power Link Family

Jiang et al. (2013) proposed a general class of power link functions, allowing flexible skewness in both positive and negative directions, while retaining the symmetric baseline link function as a special case. Let F_0^{-1} be a symmetric baseline link function with corresponding cumulative distribution function (cdf) F_0, then symmetric power link family is defined as

$$F_{sp}(x; r) = F_0^r(\frac{x}{r}) I_{(0,1]}(r) + \big[1 - F_0^{1/r}(-rx)\big] I_{(1,+\infty)}(r), \tag{2.3}$$

where $I_A(r)$ is an indicator function with 1, when $r \in A$ and 0, otherwise. $x \in (-\infty, +\infty)$ denotes covariate and $r \in (0, +\infty)$ represents a power parameter. Particularly, F_{sp} becomes

to be $F_{splogit}$ when F_0 follows a logistic distribution with location=0 and scale=1, and becomes to be F_{spep} when F_0 follows a laplace distribution location=0 and scale=1, as shown in Equation 2.4.

$$F_0 = \begin{cases} F_{logistic}(x|\mu = 0, s = 1) = \frac{1}{1+exp(-x)} \\ F_{laplace}(x|\mu = 0, b = 1) = \frac{exp(x)}{2} I_{(-\infty,0)}(x) + \big(1 - \frac{exp(x)}{2}\big) I_{[0,+\infty)}(x) \end{cases} \tag{2.4}$$

Figure 2.1 illustrates the skewness of response probability given symmetric covariate x. Particularly, splogit reduces to be the logit link when power parameter $r = 1$ and we observe that spep link provides flexible range of skewness and adjustment of tail behavior as addressed in Jiang et al. (2013).

2.2.3 Conditional Independence

Let $\boldsymbol{\theta}$ denote a vector of all unknown parameters, we have following expressions by the assumption of conditional independence:

The longitudinal continuous process is conditionally independent of the binary process,

$$p(\boldsymbol{Y}_{li}, \boldsymbol{R}_{li}|b_l, U_{li}, \boldsymbol{W}_{li}, \boldsymbol{\theta}) = p_1(\boldsymbol{Y}_{li}|b_l, U_{li}, \boldsymbol{W}_{li}, \boldsymbol{\theta}) p_2(\boldsymbol{R}_{li}|b_l, U_{li}, \boldsymbol{\theta}) \tag{2.5}$$

Repeated measurements in the longitudinal process are independent of each other,

$$p(\boldsymbol{Y}|\boldsymbol{b}, \boldsymbol{U}, \boldsymbol{W}, \boldsymbol{\theta}) = \prod_{l=1}^{L} \prod_{i=1}^{n_l} p(\boldsymbol{Y}_{li}|b_l, U_{li}, \boldsymbol{W}_{li}, \boldsymbol{\theta}) \tag{2.6}$$

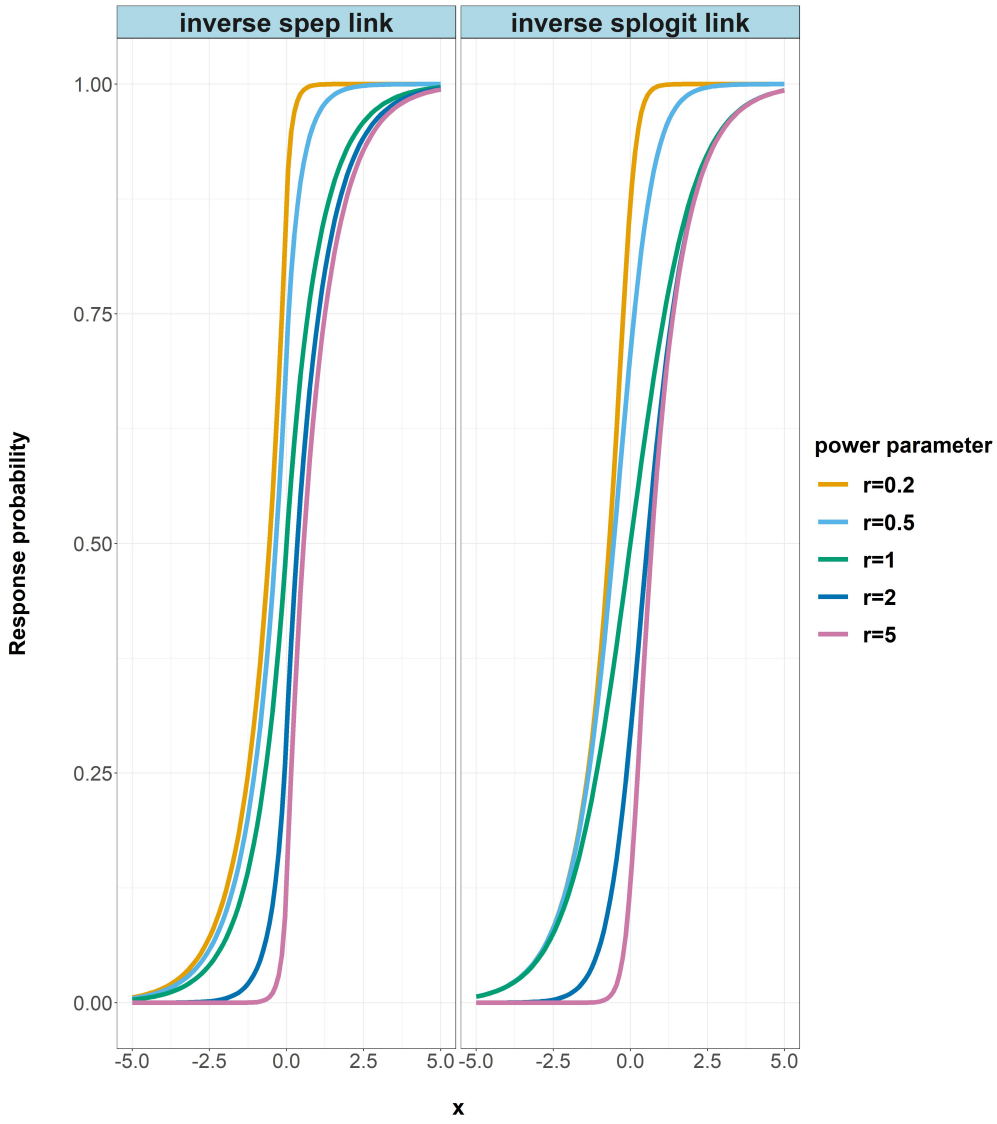

Figure 2.1: $F_{splogit}$ vs. F_{spep}

Repeated outcomes in the longitudinal binary process are independent of each other,

$$p(\boldsymbol{R}|\boldsymbol{b},\boldsymbol{U},\boldsymbol{\theta}) = \prod_{l=1}^{L}\prod_{i=1}^{n_l} p(\boldsymbol{R}_{li}|b_l,U_{li},\boldsymbol{\theta}) \tag{2.7}$$

2.2.4 Bayesian Inference

2.2.4.1 Posterior distribution

The joint posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{U}, \boldsymbol{W})$ is written as,

$$\begin{aligned}
\pi(\boldsymbol{\Psi}, \boldsymbol{b}, \boldsymbol{U}, \boldsymbol{W}|\mathcal{D}) &\propto \pi(\mathcal{D}, \boldsymbol{b}, \boldsymbol{U}, \boldsymbol{W}|\boldsymbol{\Psi})\pi(\boldsymbol{\Psi}) \\
&\propto \pi(\mathcal{D}|\boldsymbol{b}, \boldsymbol{U}, \boldsymbol{W}, \boldsymbol{\Psi})\pi(\boldsymbol{b}|\boldsymbol{U}, \boldsymbol{W}, \boldsymbol{\Psi})\pi(\boldsymbol{U}|\boldsymbol{W}, \boldsymbol{\Psi})\pi(\boldsymbol{W}|\boldsymbol{\Psi})\pi(\boldsymbol{\Psi}) \\
&\propto \pi(\boldsymbol{Y}, \boldsymbol{R}|\boldsymbol{b}, \boldsymbol{U}, \boldsymbol{W}, \boldsymbol{\Psi})\pi(\boldsymbol{b}|\boldsymbol{\Psi})\pi(\boldsymbol{U}|\boldsymbol{\Psi})\pi(\boldsymbol{W}|\boldsymbol{\Psi})\pi(\boldsymbol{\Psi}) \\
&\propto \prod_{l=1}^{L}\prod_{i=1}^{n_l} I(\boldsymbol{Y}_{li}, \boldsymbol{R}_{li}|b_l, U_{li}, \boldsymbol{W}_{li}, \boldsymbol{\Psi})\pi(b_l|\sigma_b)\pi(U_{li}|\sigma_u)\pi(\boldsymbol{W}_{li}|\tau, \rho)\pi(\boldsymbol{\Psi})
\end{aligned} \tag{2.8}$$

where $\mathcal{D}$ denotes observed data, $\boldsymbol{\Psi} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \boldsymbol{r}^T, \sigma_b, \sigma_u, \sigma, \tau, \rho, \rho_1, \rho_2)$ represents unknown mutually independent parameters and $\boldsymbol{I}(\boldsymbol{Y}_{li}, \boldsymbol{R}_{li}|b_l, U_{li}, \boldsymbol{W}_{li}, \boldsymbol{\Psi}) = \boldsymbol{I}_1(\boldsymbol{Y}_{li}|b_l, U_{li}, \boldsymbol{W}_{li}, \boldsymbol{\Psi})\boldsymbol{I}_2(\boldsymbol{R}_{li}|b_l, U_{li}, \boldsymbol{\Psi})$ by Equation 2.5.

2.2.4.2 Prior specifications

The setups of prior in Table 2.1 are based on the best practices from previous studies. Let $N(\mu, \sigma^2)$ denote normal distribution with location $\mu \in \mathbb{R}$ and scale $\sigma \in \mathbb{R}^+$; $t(\nu, \mu, \sigma)$ denote student's t distribution with degrees of freedom $\nu \in \mathbb{R}^+$, location $\mu \in \mathbb{R}$ and scale $\sigma \in \mathbb{R}^+$; Cauchy(x_0, γ) denote Cauchy distribution with location $x_0 \in \mathbb{R}$ and scale $\gamma \in \mathbb{R}^+$; Uniform(a, b) denote Uniform distribution with $-\infty < a < b < +\infty$; Gamma$(\alpha, \beta)$ denote Gamma distribution with shape $\alpha \in \mathbb{R}^+$ and rate $\beta \in \mathbb{R}^+$, of which Exponential(β) is a special case of the Gamma distribution when $\alpha = 1$. Inv-Gamma(α, β) denote Inverse-

gamma distribution with shape $\alpha \in \mathbb{R}^+$ and scale $\beta \in \mathbb{R}^+$. When the domain of Cauchy and Normal distribution is restricted (here, above zero), we name them half and truncated, respectively. The *sd* represents standard deviation of longitudinal residuals.

Table 2.1: Prior distributions

Priors	Simulation Study A	Simulation Study B	Motivating Data
$\boldsymbol{\alpha}, \boldsymbol{\beta}$	$N(0, 10^2\boldsymbol{I})$	$N(0, 100^2\boldsymbol{I})$	$N(0, 100^2\boldsymbol{I})$
σ_b	Half-Cauchy$(0, \text{scale} = 5)$	Half-Cauchy$(0, \text{scale} = 5)$	Half-Cauchy$(0, \text{scale} = 5)$
σ_u	$t(1, 0, 5)$	Half-Cauchy(0,scale=5)	Half-Cauchy(0,scale=5)
σ	Half-Cauchy$(0, \text{scale} = 2.5 \cdot sd)$	Half-Cauchy$(0, \text{scale} = 2.5 \cdot sd)$	Half-Cauchy$(0, \text{scale} = 2.5 \cdot sd)$
ρ_1, ρ_2	Uniform$(-1, 1)$	Uniform$(-1, 1)$	Uniform$(-1, 1)$
$\boldsymbol{r}$	Exponential(rate $= 1$)	Gamma(shape $= 2$, rate $= 2$)	Exponential(rate $= 1$)
τ	-	Truncated $N(0, 5^2)$	Truncated $N(0, 5^2)$
ρ	-	Inv-Gamma(shape $= 2$, scale $= 1$)	Inv-Gamma(shape $= 2$, scale $= 1$)

2.2.5 Dynamic Individual Prediction

We employ the best linear unbiased predictor (BLUP) to predict individual random intercept $U_{li'}$ and time-continuous random variable $\boldsymbol{W}_{li'}(t)$ for any new patient i' from the existing center l. The explicit forms are obtained by the properties of conditional multivariate normal distribution as utilized in Diggle et al. (2015) as

$$E\big(U_{li'}|\boldsymbol{Y}_{li'}, b_l, \boldsymbol{\phi}, \boldsymbol{\alpha}\big) = \sigma_u^2 \boldsymbol{K}_{li'}^T \big(\boldsymbol{V}_{li'}(\boldsymbol{\phi})\big)^{-1} (\boldsymbol{Y}_{li'} - \boldsymbol{X}_{li'}\boldsymbol{\alpha} - b_l \boldsymbol{K}_{li'}) \tag{2.9}$$

$$E\big(W_{li'}(t_{li'j})|\boldsymbol{Y}_{li'}, b_l, \boldsymbol{\phi}, \boldsymbol{\alpha}\big) = \tau^2 \boldsymbol{F}_{li'}^{j^T} \big(\boldsymbol{V}_{li'}^j(\boldsymbol{\phi})\big)^{-1} (\boldsymbol{Y}_{li'}^j - \boldsymbol{X}_{li'}^j\boldsymbol{\alpha} - b_l \boldsymbol{K}_{li'}^j) \tag{2.10}$$

where $\boldsymbol{Y}_{li'}$ denotes observed data and $\boldsymbol{X}_{li'}$ is the design matrix. $\boldsymbol{K}_{li'}$ represents an $n_{li'} \times 1$ matrix of ones and $\boldsymbol{\alpha}$ is as before. Also,

$$\boldsymbol{V}_{li'}(\boldsymbol{\phi}) = \sigma_u^2 \boldsymbol{J}_{li'} + \tau^2 \boldsymbol{R}_{li'} + \sigma^2 \boldsymbol{I}_{li'} \tag{2.11}$$

where $\boldsymbol{\phi} = (\sigma_2^2, \tau^2, \sigma^2, \rho)$, $\boldsymbol{J}_{li'}$ is an $n_{li'} \times n_{li'}$ all ones matrix, $\boldsymbol{R}_{li'}$ is an $n_{li'} \times n_{li'}$ matrix with $(t, t')th$ element $exp(-|t - t'| \cdot \rho)$ and $\boldsymbol{I}_{li'}$ is an $n_{li'} \times n_{li'}$ identity matrix.

Let response and covariates history up to an observed time point j be presented by $\boldsymbol{Y}_{li'}^j$ and $\boldsymbol{X}_{li'}^j$, respectively. $\boldsymbol{F}_{li'}^j = \Big(exp(-|t_{li'1} - t_{li'j}| \cdot \rho), \ldots, exp(-|t_{li'j} - t_{li'j}| \cdot \rho)\Big)^T$, $\boldsymbol{V}_{li'}^j$ is the variance-covariance matrix of $\boldsymbol{Y}_{li'}^j$. Analogously, forecasting $W_{lij'}$ at time t_{lij} with lead-time u for an existing patient i is obtained by

$$E\big(W_{li}(t_{lij} + u) | \boldsymbol{Y}_{li}^j, b_l, U_{li}, \boldsymbol{\phi}, \boldsymbol{\alpha}\big) = \tau^2 \boldsymbol{F}_{li}^{j,u^T} \big(\boldsymbol{G}_{li}^j(\boldsymbol{\phi})\big)^{-1} (\boldsymbol{Y}_{li}^j - \boldsymbol{X}_{li}^j \boldsymbol{\alpha} - b_l \boldsymbol{K}_{li}^j - U_{li} \boldsymbol{K}_{li}^j) \tag{2.12}$$

where $\boldsymbol{G}_{li}^j(\boldsymbol{\phi}) = \tau^2 \boldsymbol{R}_{li}^j + \sigma^2 \boldsymbol{I}_{li}^j$, $\boldsymbol{R}_{li}^j$ and $\boldsymbol{I}_{li}^j$ are as before but with new dimensions of $j \times j$. $\boldsymbol{F}_{li}^{j,u} = \Big(exp(-|t_{lij} + u - t_{li1}| \cdot \rho), \ldots, exp(-|t_{lij} + u - t_{lij}| \cdot \rho)\Big)^T$. Whenever a new response at time $t_{lij'}$ becomes available, the prediction from Equation 2.12 will be dynamically updated. Practically, we replace all unknown parameters with the posterior mean estimates to achieve the empirical BLUP.

2.2.6 Model Selection

In this section we consider a recent measure of comparison between models. The Watanabe-Akaike information criterion (WAIC, Watanabe (2010)) is favored as a full Bayesian ap-

proach for estimating the out-of-sample expectation, by marginalizing the posterior distribution rather than conditioning on a point estimate, thus is invariant to reparameterisations contrary to the Deviance Information Criterion (DIC, Spiegelhalter et al. (2002)). In a comprehensive review paper, Gelman and his colleagues (Gelman et al. (2014)) recommended expression on the deviance scale in terms of variance adjustment. Thus computed WAIC_1 and WAIC_2 for the two submodels are written as follows,

$$\widehat{\text{WAIC}}_1 = -2\Big\{ \sum_{l=1}^{L} \sum_{i=1}^{n_l} \sum_{j=1}^{n_{li}} \Big(log\big[\frac{1}{S} \sum_{s=1}^{S} \boldsymbol{I}_1(Y_{lij}|b_l^s, U_{li}^s, W_{lij}^s, \boldsymbol{\Psi}^s) \big] - V_{s=1}^S log\big[\boldsymbol{I}_1(Y_{lij}|b_l^s, U_{li}^s, W_{lij}^s, \boldsymbol{\Psi}^s) \big] \Big) \Big\} \tag{2.13}$$

$$\widehat{\text{WAIC}}_2 = -2\Big\{ \sum_{l=1}^{L} \sum_{i=1}^{n_l} \sum_{j=1}^{n_{li}} \Big(log\big[\frac{1}{S} \sum_{s=1}^{S} \boldsymbol{I}_2(R_{lij}|b_l^s, U_{li}^s, \boldsymbol{\Psi}^s) \big] - V_{s=1}^S log\big[\boldsymbol{I}_2(R_{lij}|b_l^s, U_{li}^s, \boldsymbol{\Psi}^s) \big] \Big) \Big\} \tag{2.14}$$

where $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ denote the likelihood function for normal distribution and bernoulli distribution, respectively. S is the number of simulation draws, $\boldsymbol{\Psi}^s$ is the vector of the model parameters at s^{th} iteration. $V_{s=1}^S$ represents the sample variance, that is $V_{s=1}^S \boldsymbol{a} = \frac{1}{S-1} \Sigma_{s=1}^S (a_s - \bar{a})^2$. Practically, calculations on Equation 2.13 and Equation 2.14 could be easily achieved by the waic function from the loo package (v2.3.1) (Vehtari et al. (2020)) by extracting pointwise log-likelihood values from the posterior samplings. The model with the smaller WAIC indicates the better goodness of fit.

2.3 Simulated Studies

2.3.1 Simulation Study 1

The first simulation study is designed to explore the flexibility of splogit in a comparison with three common links, namely logit, probit, cloglog. Hereafter, we name them by the rule of link-JM. The explicit expressions are shown as

$$\begin{cases} g_{logit}^{-1}(x) = F_{logit}(x) = [1 + exp(x)]^{-1} \\ g_{probit}^{-1}(x) = F_{probit}(x) = \Phi(x), \text{where } \Phi \text{ is standard normal cdf} \\ g_{cloglog}^{-1}(x) = F_{cloglog}(x) = 1 - e^{-e^x} \\ g_{splogit}^{-1}(x;r) = F_{splogit}(x;r) = [1 + exp(\frac{x}{r})]^{-r} I_{(0,1]}(r) + \left\{1 - [1 + exp(-rx)]^{-\frac{1}{r}}\right\} I_{(1,+\infty)}(r) \end{cases} \tag{2.15}$$

where $x \in (-\infty, +\infty)$ denotes covariate, $r \in (0, +\infty)$ represents the power parameter. We simulate multicenter data through proposed splogit-JM to fit following joint models, of the forms,

$$
\text{logit-JM}\begin{cases} Y_{lij} = \alpha_0 + x_{li1}\alpha_1 + x_{li2}\alpha_2 + b_l + U_{li} + \epsilon_{lij} \\ \Pr(R_{lij}=1) = F_{logit}(\beta_0 + \beta_1 t_{lij} + \rho_1 b_l + \rho_2 U_{li}) \end{cases}
$$

$$
\text{probit-JM}\begin{cases} Y_{lij} = \alpha_0 + x_{li1}\alpha_1 + x_{li2}\alpha_1 + b_l + U_{li} + \epsilon_{lij} \\ \Pr(R_{lij}=1) = F_{probit}(\beta_0 + \beta_1 t_{lij} + \rho_1 b_l + \rho_2 U_{li}) \end{cases}
$$

$$
\text{cloglog-JM}\begin{cases} Y_{lij} = \alpha_0 + x_{li1}\alpha_1 + x_{li2}\alpha_1 + b_l + U_{li} + \epsilon_{lij} \\ \Pr(R_{lij}=1) = F_{cloglog}(\beta_0 + \beta_1 t_{lij} + \rho_1 b_l + \rho_2 U_{li}) \end{cases}
$$

$$
\text{splogit-JM}\begin{cases} Y_{lij} = \alpha_0 + x_{li1}\alpha_1 + x_{li2}\alpha_1 + b_l + U_{li} + \epsilon_{lij} \\ \Pr(R_{lij}=1) = F_{splogit}(\beta_0 + \beta_1 t_{lij} + \rho_1 b_l + \rho_2 U_{li}; r_l) \end{cases}
$$

where all notations are as before. Without loss of generality, we simulate a balanced data set, that is $L = 5, n_l = 50, n_{li} = 5$ and set regression coefficients $\alpha_0 = 1, \alpha_1 = -3, \alpha_2 = 0.7, \beta_0 = 0.5, \beta_1 = 4$, power parameter $\boldsymbol{r} = (0.2, 0.5, 1, 2, 3)^T$, association parameters $\rho_1 = 0.5, \rho_2 = 0.8$. Let $x_{li1} \sim N(0,1), x_{li2} \sim$ Bernoulli(0.5), $b_l \sim N(0, \sigma_b^2), U_{li} \sim N(0, \sigma_u^2)$ and $\epsilon_{lij} \sim N(0, \sigma^2)$, with $\sigma_b = 5, \sigma_u = 3, \sigma = 0.45$. Time point t_{lij} is simulated from a uniform(0,6) distribution by assuming every patient starts from time=0 and is standardized to ensure the computational stability.

All models are estimated by HMC with 4000 post-warmup iterations via two chains for 50 replicates. Averaged simulation results and WAIC are summarized in Figure 2.2 and Table 2.2, respectively. Posterior samplings are ensured to converge until all values for the potential scale reduction factor $\hat{R}$ are below 1.1 (Gelman and Rubin (1992)).

Table 2.2: Model comparisons over simulated data sets for 50 replicates

Fitted Model	WAIC [a]	$WAIC_1$ [b]	$WAIC_2$ [c]
logit-JM	2239.93	1807.16	432.77
probit-JM	2244.44	1807.49	436.95
cloglog-JM	2272.57	1808.24	464.33
splogit-JM	**2203.00**	**1806.37**	**396.63**

Model in boldface: true model; [a] Joint model; [b] Longitudinal continuous submodel; [c] Longitudinal binary submodel

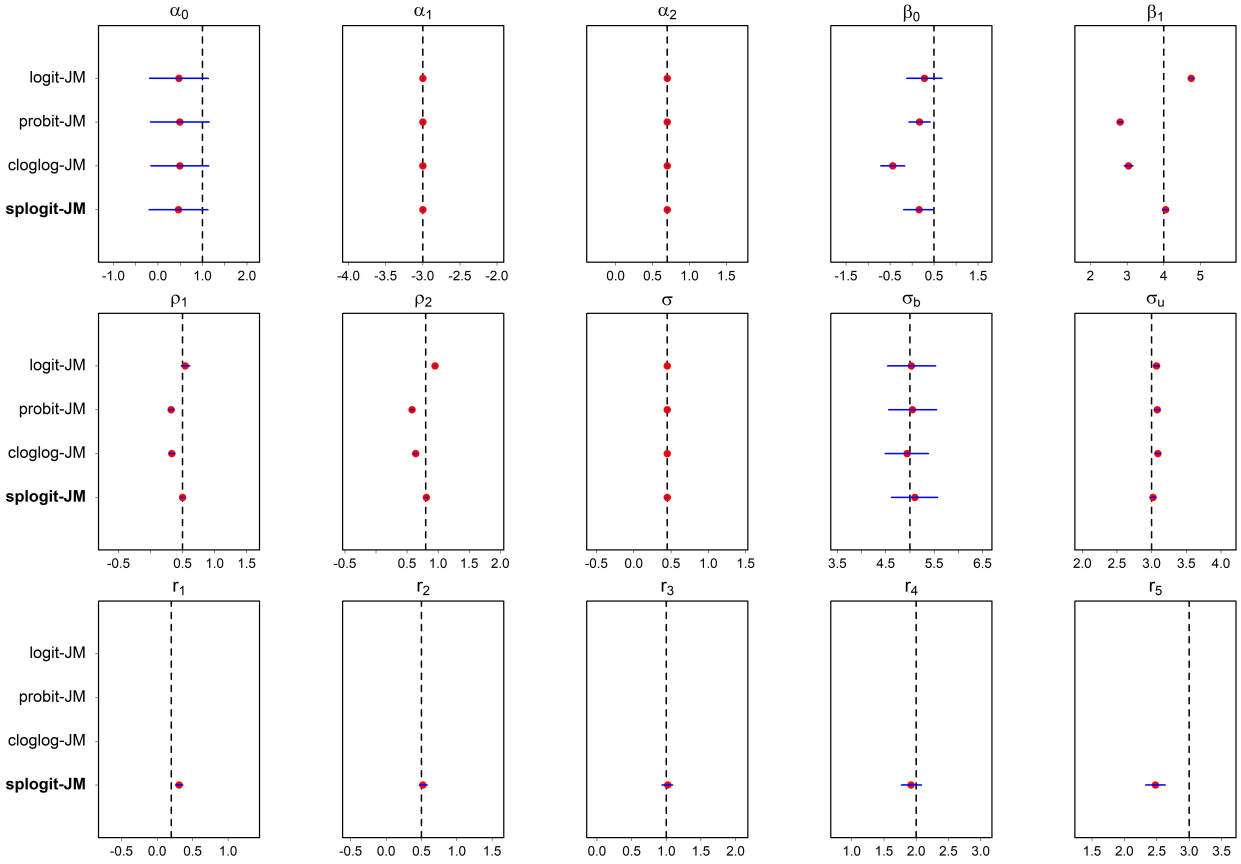

Figure 2.2: Averaged posterior mean (red dot) with true model (boldface), true value (dashed line) and 95% confidence interval (blue line) for 50 replicates via CmdStanr.

The smallest WAIC_2 from Table 2.2 demonstrates that center-specific power parameter induces great flexibility than common fixed links in a hierarchical data structure. Among those conventional links, logit-JM as the special case of true splogit-JM outperforms others. The visualization from Figure 2.2 shows the reasonable performance from splogit-JM in recovering the true values (e.g., β_1) and we concur that estimates for α_0 and r_5 can be further improved by more replicates. To summarize, a flexible link function is preferred for multicenter data from our simulation results.

2.3.2 Simulation Study 2

As we see from previous simulation, a flexible link function is recommended over a fixed link function. Given the advantages of spep compared to the splogit (see Figure 2.1), we choose spep as the flexible link function for the following analysis. In addition. we take the LME submodel without fixed effects for simplicity purpose and JM equations are specified as follows

$$
\text{spep-JM}_1 \begin{cases} Y_{lij} = U_{li} + \epsilon_{lij} \\ \Pr(R_{lij} = 1) = F_{spep}(\beta_0 + \beta_1 t_{lij} + \rho_2 U_{li}; r) \end{cases}
$$

$$
\text{spep-JM}_2 \begin{cases} Y_{lij} = b_l + U_{li} + \epsilon_{lij} \\ \Pr(R_{lij} = 1) = F_{spep}(\beta_0 + \beta_1 t_{lij} + \rho_1 b_l + \rho_2 U_{li}; r) \end{cases}
$$

$$
\text{spep-JM}_3 \begin{cases} Y_{lij} = b_l + U_{li} + \epsilon_{lij} \\ \Pr(R_{lij} = 1) = F_{spep}(\beta_0 + \beta_1 t_{lij} + \rho_1 b_l + \rho_2 U_{li}; r_l) \end{cases}
$$

$$
\text{spep-JM}_4 \begin{cases} Y_{lij} = b_l + U_{li} + W_{lij} + \epsilon_{lij} \\ \Pr(R_{lij} = 1) = F_{spep}(\beta_0 + \beta_1 t_{lij} + \rho_1 b_l + \rho_2 U_{li}; r_l) \end{cases}
$$

where spep-JM_1 is constructed as a non-hierarchical joint model and spep-JM_4 is the proposed joint model with exponential correlation matrix and center-specific power parameter; t_{lij} is a standardized age variable for the stabilization of posterior computations. We set regression coefficients $\beta_0 = 0.5$, $\beta_1 = 4$ and association parameters $\rho_1 = 0.5$, $\rho_2 = 0.8$. Let $b_l \sim N(0, \sigma_b^2)$, $U_{li} \sim N(0, \sigma_u^2)$ and $\epsilon_{lij} \sim N(0, \sigma^2)$, with $l = 1, \ldots, 5, i = 1, \ldots, 50, j = 1, \ldots, 10, \sigma_b = 5, \sigma_u = 3, \sigma = 0.45$; $\boldsymbol{W}_{li}(t)$ be the stationary Gaussian process with $\tau = 1.5$ and $\rho = 0.5$ as defined in Section 2.2. We simulate a total of 50 data sets under true power parameters $\boldsymbol{r} = (0.2, 0.5, 1, 1, 2)^T$. Priors are defined according to Section 2.2.4. All models are estimated by HMC with at least 10,000 iterations via two chains. The first 4000 draws are discarded as a warm-up sampling and every third values are kept for the posterior inference. We ensure all results satisfy the convergence diagnostic $\hat{R}$.

Table 2.3: Model comparisons over simulated data sets for 50 replicates

Fitted Model	WAIC[a]	$WAIC_1$[b]	$WAIC_2$[c]
Non-hierarchical (spep-JM_1)	9256.99	8299.09	957.90
Common power (spep-JM_2)	9179.92	8292.14	887.78
Naive LME (spep-JM_3)	9157.46	8291.87	865.59
Proposed (spep-JM_4)	**8352.83**	**7541.87**	**810.96**

Model in boldface: true model; [a] Joint model; [b] Longitudinal continuous submodel; [c] Longitudinal binary submodel

As shown in Table 2.3, spep-JM_4 is correctly identified by the lowest WAIC, while spep-JM_1 performs the worst, reflecting that ignoring center effect in a multicenter cohort study yields poor performance. spep-JM_3 performs better than spep-JM_2, indicating that center-specific power parameter is preferred. Figure 2.3 illustrates that posterior mean estimates of the true joint model have negligible bias, indicating that the proposed joint model provides valid Bayesian inference. The striking biases from all other three joint models are found from scale parameter (σ) for the measurement error. In addition, scale parameter for patients (σ_u) produced from non-hierarchical spep-JM_1 is misleading compared to the true value.

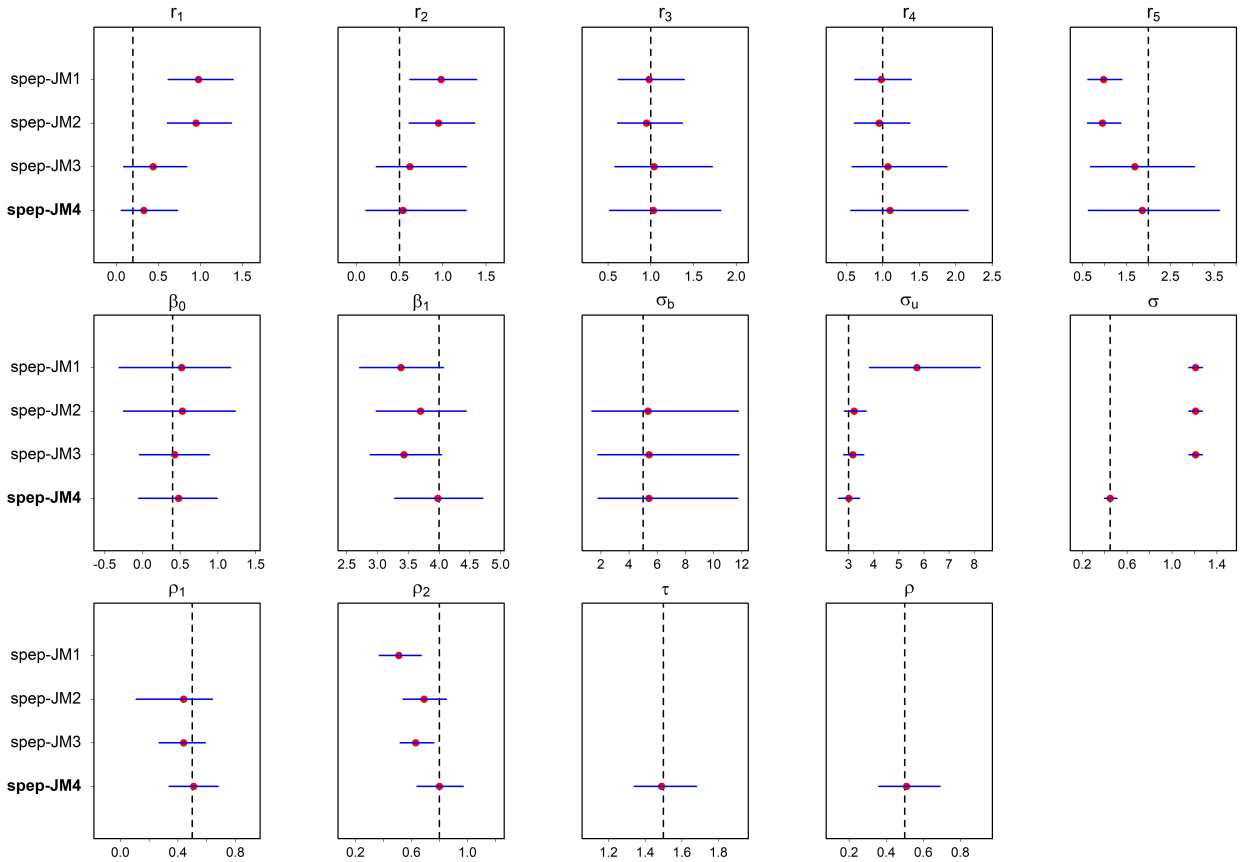

Figure 2.3: Averaged posterior means by 50 replicates via RStan with true model (boldface), rue value (dashed vertical line), posterior mean estimates (red dot) and corresponding 95% credible interval (blue line). JM1: Misspecified; JM2: No center-index; JM3: No covariance.

2.4 Application of CF Study

2.4.1 Motivating Data

Our analysis cohort for this application consists of 381 CF patients who contributed a total of 9,209 observations across five centers (see more details in Appendix B.2). These centers are randomly selected among those with feasible sample size between 50 and 100 for computational aspects of the modeling. Around 15% of measurements are excluded due to missing

values on ppFEV1 and Body Mass Index (BMI). We observe neither drop-outs (e.g. death or lung transplantation) nor patients switch centers during the observed period. Data cleaning and descriptive statistics are summarized in Appendix B. Approval for the data analysis was made by the Institutional Review Board (IRB) from Cincinnati Children's Hospital Medical Center (CCHMC) and University of Cincinnati (UC).

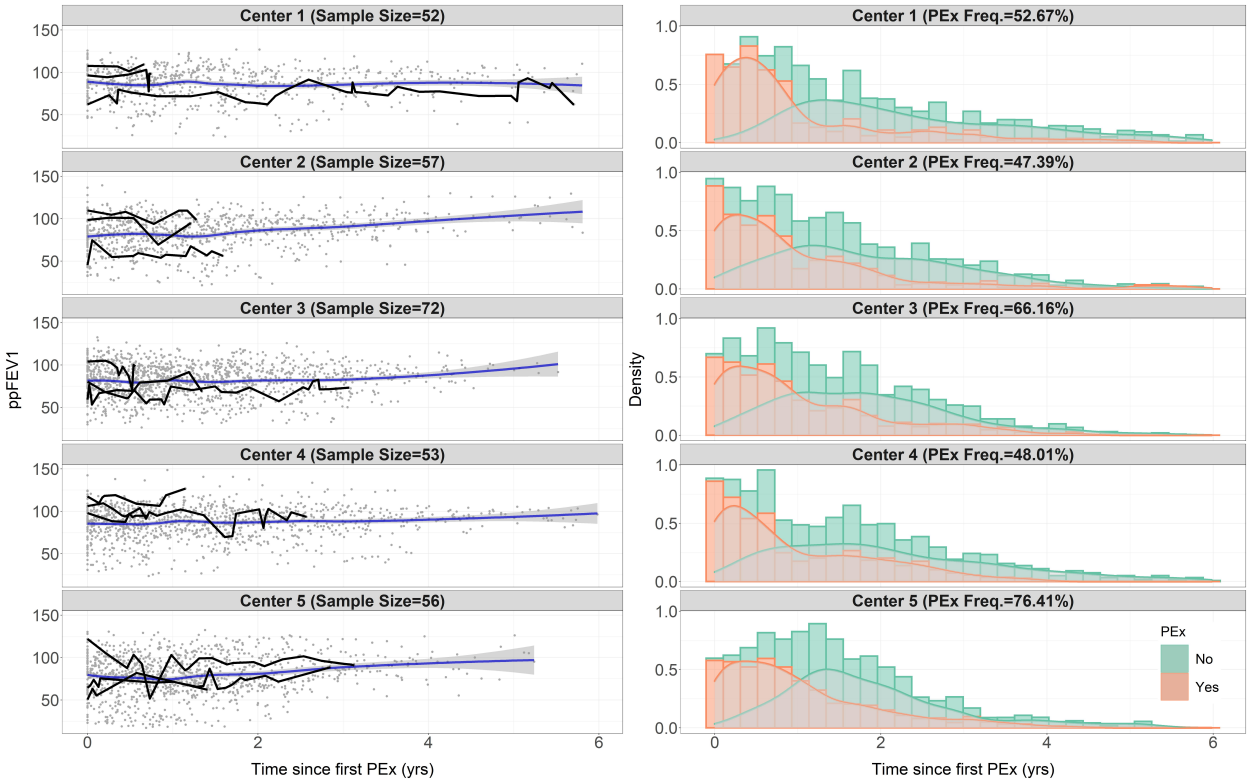

Figure 2.4: Observed ppFEV1 (left panel) and density of PEx (right panel) against time since the first PEx occurrence in years. Within each center including: three random profiles (black lines), observed values (gray dots) and LOWESS smoothing curves with 95% confidence interval (blue lines with gray-shaded bands); histograms (bars) with densities (areas) grouped by PEx occurrence; Acronym: Freq. = Frequency

We restrict encounter age to include observations taken under valid pulmonary function testing (e.g., above 6 years old) and up to early adolescence (e.g., 12 years old). The upper

limit is chosen to restrict the cohort to the first bout of lung function decline, which is shown to occur before more rapid decline during late adolescence and early adulthood (Szczesniak et al. (2013)). To be included in the analysis, patients should have at least three or more observed ppFEV1 measurements spanning at least six months after the first PEx occurrence. The time period selected for this analysis is from calendar year 2003 to 2017, because modern predictors with encounter levels are consistently documented beginning in 2003 (Knapp et al. (2016)). Average (range) duration of follow-up is 4.08 (0.53-5.99) years. Figure 2.4 displays the trajectory of ppFEV1 and density of PEx throughout the whole analysis period. The left panel demonstrates the heterogeneous nature of ppFEV1 with a slightly increasing, smoothing curve in each center. The right panel indicates some underlying skewness of PEx responses.

2.4.2 Internal Validation

To access the predictive performance, we conduct 80-20% cross-validation by choosing a random sample of 20% of patients from each center, in which 79 patients with 1993 observations are for testing cohort. Among those remaining 80% patients, we select the first 80% observations with at least three records spanning at least six months after the first PEx for the training cohort, which consists of 302 patients with 5769 observations and the rest of their 1447 observations are conformed as the masking cohort.

2.4.3 Parameter Estimation

Predictors are selected by a conventional Stepwise method (Hocking (1976)) from a two-stage basis as suggested in Barrett et al. (2019). We apply the R function buildlme from buildmer package (Voeten (2021)) for LME submodel and the R function step from stats package (R Core Team (2020)) for GLMM. By defining the baseline as the time of the first PEx event, we finally include baseline ppFEV1, BMI percentile, Methicillin-resistant Staphylococcus aureus (MRSA), pseudomonas aeruginosa (pa) and genotype F508del heterozygote for the LME submodel; time since baseline (in years), BMI percentile, pancreatic enzyme usage and pa for the GLMM.

Table 2.4: Model comparisons for CF data with the boldface as the optimal model.

Fitted Model	WAIC[a]	WAIC_1[b]	WAIC_2[c]
Non-hierarchical (spep-JM_1)	49960.1	44302.0	5658.1
Common power (spep-JM_2)	49867.6	44244.8	5622.8
Naive LME (spep-JM_3)	49812.0	44228.0	5584.0
Proposed (spep-JM_4)	**47082.3**	**42704.8**	**4377.5**

[a] Joint model; [b] Longitudinal continuous submodel; [c] Longitudinal binary submodel

We examine the four joint models as constructed in Section 2.3.2. Posterior samplings are carried out by HMC with 2000 post-warmup draws via two chains and diagnostics plots (see Appendix B.3) show that all samplings are well converged. Model performance with respect to WAIC is presented in Table 2.4, undoubtedly model spep-JM_4 outperforms others with the lowest WAIC. Nonetheless, spep-JM_1 yields the worst WAIC, demonstrating

underlying biases caused by a non-hierarchical model in our CF study. To further evaluate the assumption of longitudinal continuous submodel in spep-JM$_4$, we have plotted residual diagnostics in Appendix B.4. No striking violations are found by the visual inspections.

Table 2.5: Model estimations under spep-JM$_4$

	mean [1]	se [2]	sd [3]	2.5% [4]	97.5% [4]	n_eff [5]	Rhat [6]
Longitudinal continuous submodel							
α_1 (intercept at age 6 years)	27.64	0.07	2.50	22.68	32.75	1333	1.00
α_2 (BMI percentile)	0.20	0.00	0.01	0.18	0.22	1960	1.00
α_3 (ppFEV1 at baseline)	0.56	0.00	0.03	0.51	0.61	1273	1.00
α_4 (MRSA)	-1.02	0.01	0.50	-1.99	-0.06	2265	1.00
α_5 (pa)	-0.54	0.01	0.42	-1.38	0.27	2191	1.00
α_6 (F508del Heterozygote)	2.47	0.03	1.23	-0.04	4.83	1284	1.00
σ_b (between centers, intercept)	1.49	0.05	1.26	0.07	4.51	702	1.00
σ_u (between patients, intercept)	3.38	0.02	0.66	2.12	4.70	716	1.00
σ (measurement error)	8.77	0.00	0.11	8.55	8.98	1563	1.00
τ (scale parameter)	11.20	0.01	0.36	10.53	11.93	1463	1.00
ρ (1/range)	0.41	0.00	0.04	0.34	0.50	1213	1.00
Longitudinal binary submodel							
β_1 (intercept at age 6 years)	2.94	0.01	0.23	2.53	3.40	1375	1.00
β_2 (time)	-0.37	0.00	0.02	-0.42	-0.33	1921	1.00
β_3 (BMI percentile)	-0.01	0.00	0.00	-0.01	-0.01	1535	1.00
β_4 (Enzymes)	-0.13	0.00	0.06	-0.25	-0.03	1812	1.00
β_5 (pa)	-0.22	0.00	0.07	-0.35	-0.09	2030	1.00
r_1 (power parameter, Center 1)	3.99	0.03	1.41	1.94	7.42	1888	1.00
r_2 (power parameter, Center 2)	2.76	0.03	1.24	1.27	6.15	1509	1.00
r_3 (power parameter, Center 3)	2.76	0.03	1.12	1.29	5.62	1673	1.00
r_4 (power parameter, Center 4)	4.23	0.04	1.60	1.86	7.90	1480	1.00
r_5 (power parameter, Center 5)	2.93	0.03	1.21	1.40	6.09	1831	1.00
Association structure							
ρ_1 (submodel link, center)	-0.20	0.01	0.30	-0.81	0.63	403	1.01
ρ_2 (submodel link, patient)	-0.44	0.00	0.10	-0.68	-0.30	607	1.00

[1] posterior mean; [2] Monte Carlo standard error; [3] Monte Carlo standard deviation; [4] posterior quantiles; [5] effective sample size; [6] potential scale reduction factor (at convergence, Rhat=1)

The estimations for spep-JM_4 are summarized in Table 2.5. For longitudinal continuous submodel, baseline ppFEV1 (0.56 [0.51, 0.61]) and BMI percentile (0.2 [0.18, 0.22]) imply significantly positive relationship with ppFEV1. In Taylor-Robinson's study (Taylor-Robinson et al. (2012)), they found that people with high ppFEV1 at baseline were more likely to have a higher ppFEV1 up to 15 years, which reconciles the positive effect of baseline ppFEV1 as in ours. Categorical predictors MRSA (-1.02 [-1.99, -0.06]) and pa (-0.54 [-1.38, 0.27]) correspond to worsen overall ppFEV1, despite pa is not significant because 95% credible interval contains zero. Significant parameter ρ (0.41 [0.34, 0.5]) indicates that the correlation between two measurements within a patient decays as time elapses. For longitudinal binary submodel, the risk of PEx onset is decreasing against time (-0.37 [-0.42, -0.33]) given all the other predictors unchanged. BMI percentile (-0.01 [-0.01, -0.01]) plays an important role in ppFEV1, however, it seems not to contribute much for PEx. The infection with pa (-0.22 [-0.35, -0.09]) and pancreatic enzyme usage (-0.13 [-0.25, -0.03]) are associated with lower PEx frequency. The interpretations would be most likely that patients who are infected by pa or with insufficient pancreatic enzyme are given primary medical care. Negative association parameter ρ_1 (-0.2 [-0.81, 0.63]) suggests that the center with more severe CF patients (indicated by the lower b_l) tends to have higher risk of PEx; Analogously, negative ρ_2 (-0.44 [-0.68, -0.30]) significantly indicates that patients with lower averaged lung function (indicated by the lower U_{li}) are more likely to experience PEx.

2.4.4 Predictive Performance

We utilize root mean squared error (RMSE) and area under curve (AUC) to evaluate longitudinal continuous and binary predictions, respectively. We prefer smaller value of RMSE as it measures errors, on the contrary, AUC is a measurement for the classification, the higher AUC, the better the model is at distinguishing between PEx and non-PEx events. Corresponding results along with residuals standard error and 95% confidence intervals are summarized in Table 2.6 and Table 2.7 based on two prognostic cohorts.

Table 2.6: Predictive performance between training and testing cohorts

	Training				Testing			
	ppFEV1		PEx		ppFEV1		PEx	
	RMSE	SE	AUC	95% CIs	RMSE	SE	AUC	95% CIs
Non-hierarchical (spep-JM_1)	10.755	0.142	0.748	(0.734, 0.763)	10.397	0.233	0.623	(0.594, 0.651)
Common power (spep-JM_2)	10.686	0.141	0.748	(0.734, 0.763)	10.399	0.233	0.622	(0.593, 0.651)
Naive LME (spep-JM_3)	10.671	0.141	0.755	(0.741, 0.770)	10.398	0.233	0.639	(0.610, 0.667)
Proposed (spep-JM_4)	7.768	0.102	0.882	(0.873, 0.892)	6.879	0.154	0.631	(0.604, 0.658)

Abbreviations: RMSE=Root Mean Square Error; SE=Standard Error; AUC=Area under Curve; CI=Confidence Interval

Table 2.7: Forecasting performance between training and masking cohorts

	Training				Masking			
	ppFEV1		PEx		ppFEV1		PEx	
	RMSE	SE	AUC	95% CIs	RMSE	SE	AUC	95% CIs
Non-hierarchical (spep-JM_1)	10.755	0.142	0.748	(0.734, 0.763)	10.354	0.272	0.655	(0.626, 0.683)
Common power (spep-JM_2)	10.686	0.141	0.748	(0.734, 0.763)	10.298	0.270	0.628	(0.598, 0.658)
Naive LME (spep-JM_3)	10.671	0.141	0.755	(0.741, 0.770)	10.353	0.271	0.612	(0.582, 0.642)
Proposed (spep-JM_4)	7.768	0.102	0.882	(0.873, 0.892)	8.850	0.233	0.785	(0.760, 0.809)

Abbreviations: RMSE=Root Mean Square Error; SE=Standard Error; AUC=Area under Curve; CI=Confidence Interval

Table 2.6 shows that spep-JM_4 achieves the smallest RMSE and the highest AUC for the training cohort. We note that the difference in accuracy is not evident among JMs for the

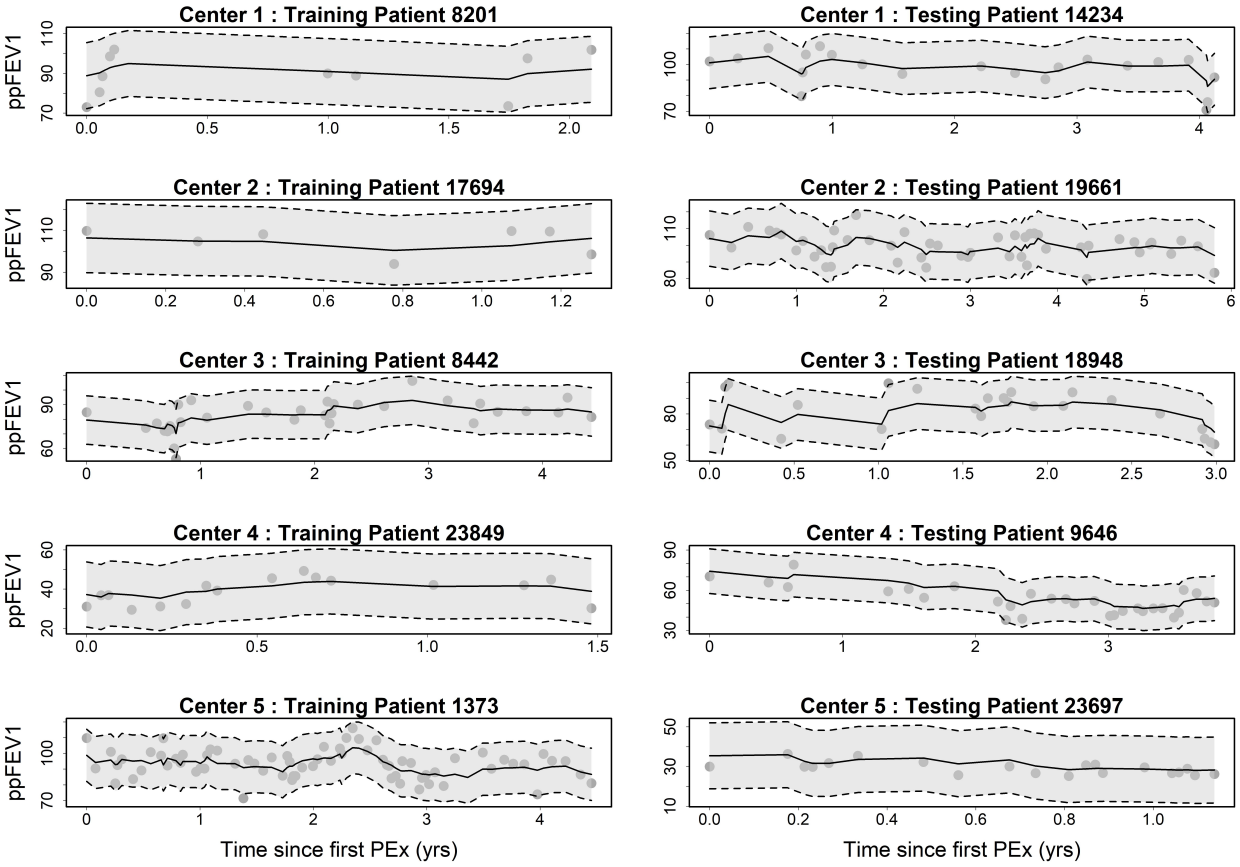

Figure 2.5: Prediction for random selected patients from each center under spep-JM_4 model, including observed ppFEV1 (gray dots) against time with fitted values (solid lines) and corresponding 95% CIs (bands)

testing cohort, however, the smallest RMSE is convincing enough to conclude spep-JM_4 as the optimal choice. We apply the same training cohort to examine the forecasting performance and Table 2.7 also demonstrates that spep-JM_4 is outstanding in forecasting performance. Figure 2.5 and Figure 2.6 present individual prediction and forecast under spep-JM_4 structure, respectively. Our proposed model is well shown to capture the heterogeneous nature of ppFEV1, whilst provide reasonable predictive probability for PEx encounters.

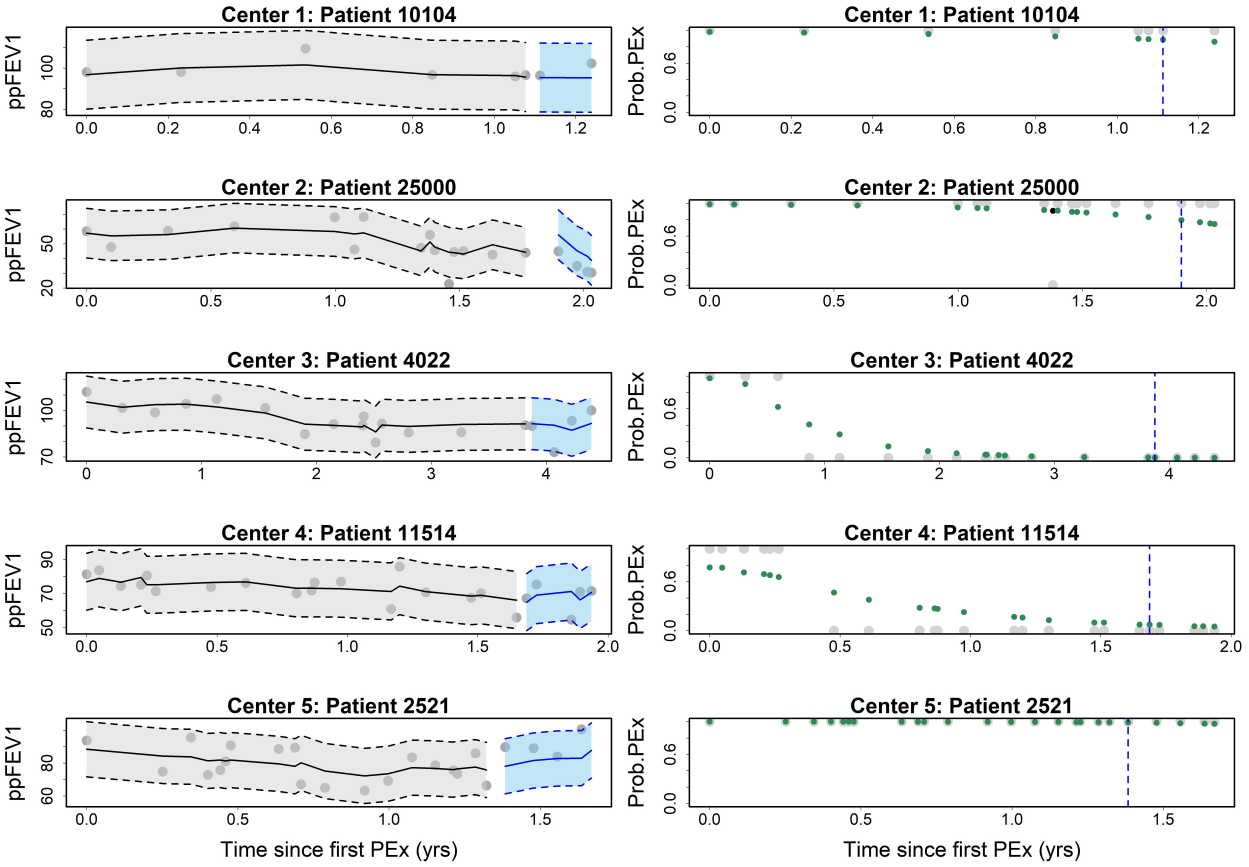

Figure 2.6: Forecast for random selected patients from each center under spep-JM_4 model, including observed ppFEV1 (gray dots) against time with fitted (solid black lines) and prognostic values (solid blue lines) and corresponding 95% CIs (bands); observed PEx (gray dots) against time with predicted probability of PEx onset (green dots for the true classification; black dots for the false classification)

2.5 Discussion

In this chapter, we have developed a multilevel Bayesian joint model with a flexible link function to accommodate analysis of longitudinal continuous and binary outcomes for a monitoring CF registry data, which is depicted by hierarchical structure and irregularly observed time points. Our novel approach relaxes the inference on regular clinical follow-up

from a previous study (Su et al. (2020), Su et al. (2021)), which avoids unnecessary biases caused by annualized covariates. The rationale for dynamic individual prediction is obtained by plugging posterior means from HMC method into BLUP equations as shown in Section 2.2.5. The interplay of Bayesian and frequentist approach is beneficial to both numerical and analytic analysis, especially in modern hierarchical models (Bayarri and Berger (2004)). Both the simulation study and motivating example demonstrate the reliable capability of our proposed joint model. Furthermore, our model can be applied to numerous alternative hierarchical data structures, see Section 2 Brilleman et al. (2019) for more data examples. The authors of the seminal flexible link function (Jiang et al. (2013)) demonstrated that, Equation 2.3 achieved left skewness when $0 < r < 1$; symmetric when $r = 1$; right skewness when $r > 1$. Researchers need to be aware that this statement cannot hold when values of covariate x are asymmetric (see the simulation study in Appendix B.1). As a conclusion, we recommend a flexible link function regardless of whether there exists observed skewness of responses in the real data case.

There may be some possible limitations in this study. We have focused our study on symmetric power link family only, however, an alternative is subject to a class of link functions based on the GEV distribution (Wang and Dey (2010)). To the best of our knowledge, there are no existing R packages that incorporate any aforementioned flexible link families, which might be an interesting field to explore. Another potential issue with our employed Stepwise algorithm has been raised in a recent paper by Harhay et al. (2020). They suggested that such procedure might lead to bias and overfitting problems. The backwards variable selection

algorithm with a large significant level seems to be a better approach (Heinze and Dunkler (2017)). From the clinical perspective in our application context, Stepwise approach provides a straightforward means of feature selection, and from the statistical point of view, predictors have minimal impacts on predictive performance as long as the covariance structure is correctly specified. The past and current work highlight the need to proceed carefully about intended purpose of the prediction model while prioritizing feature selection elements as appropriate. In addition, we may generalize the individual prediction method to account for a new patient from a new center. Given the fact that PEx onset would be dependent on various intrinsic factors and CF is a multi-system disease, the aforementioned time-independent association structure can be further extended (e.g., latent Gaussian process).

Chapter 3

Multilevel Bayesian joint model of longitudinal continuous and recurrent outcomes

In the previous chapter, we propose a multilevel joint model that accommodates longitudinal and binary outcomes. It also might be of clinical interest to monitor and predict the probability of next PEx in a survival context. In this chapter, we illustrate our proposed multilevel Bayesian joint model that encompasses the LME for longitudinal outcomes and the stratified relative risk frailty model for time-to-recurrent events.

This chapter is organized as follows. We emphasize the motivation in Section 3.1. In Section 3.2, we introduce the methodology, including framework of submodels, Bayesian

inference, predictive metrics and model selection. In Section 3.3 and Section 3.4, we illustrate a simulation study and a motivating example, respectively. Lastly, we conclude our study with remarks and discussions in Section 3.5.

3.1 Motivation

In a comprehensive review paper, Hickey et al. (2018a) summarized corresponding literature for joint models involving multivariate event time data, of which one stream is for recurrent events. To the best of our knowledge, numerous authors have studied jointly modeling longitudinal outcomes and recurrent events (either in the presence of a terminal event, or not), but not for hierarchical data structure (Liu et al. (2008), Kim et al. (2012), Musoro et al. (2015), Shen et al. (2016), Ren et al. (2021)) or vice versa (Luo and Wang (2014), Brilleman et al. (2019)). To this end, we are motivated to postulate a shared parameter joint model consisting of two submodels, the linear mixed effect (LME) submodel for the longitudinal trajectory and extended stratified relative risk frailty model (Rizopoulos (2012c)) for the recurrent event submodel. In addition, the two submodels are linked by a center-specific time-dependent latent trajectory. It is worth noting that the growing framework of latent class joint model can be employed for heterogeneous population (Han et al. (2007), Brombin et al. (2016)). However, such model assumes that the subpopulations are latent, or in other words that heterogeneity is not captured by any of the observed covariates. Consequently, the latent class joint model might not be suitable to our study due to known center information.

The visualization of our data structure along with two types of time scales to the repeated events are displayed in Figure 3.1. Calendar time and gap time are of the same interval time length, however, interpretations for risk predictions are different. With gap time, predictions are made for the next event at the end time of the previous event, while with calendar time, predictions are made for a primary event time at the study entry time (Smedinga et al. (2017)). In other words, calendar time evaluates the effect of a covariate since the study entry point, while gap time evaluates the effect since the end time of the previous event.

3.2 Model Framework

The proposed joint model with shared parameter consists of a longitudinal submodel and a time-to-recurrent event submodel, specified separately for each type of outcomes. The two submodels are linked using a latent trajectory, which is parameterised in current value and time-dependent slope of the linear predictor.

3.2.1 Longitudinal Submodel

We define $y_{lij}(t) = y_{li}(t_{lij})$ which corresponds to the observed longitudinal outcome for the i^{th} $(i = 1, \ldots, n_l)$ individual who comes from the l^{th} $(l = 1, \ldots, L)$ center taken at time point t_{lij} $(j = 1, \ldots, n_{li})$. Then the mixed effects model for $y_{lij}(t)$ takes the form

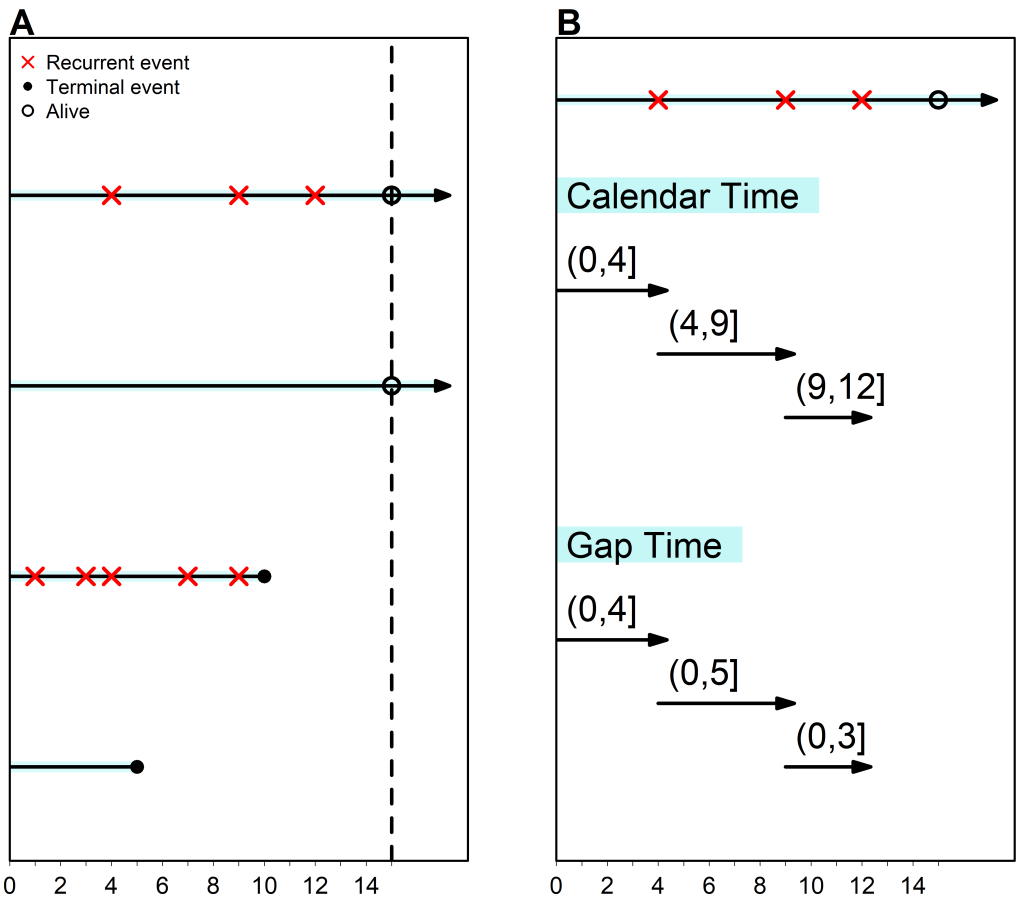

Figure 3.1: Illustrations of data structure and time scales. A: PEx occurrences for four possible patients; B: Recurrent events under two risk scenarios.

$$\begin{cases} y_{lij}(t) = m_{lij}(t) + \epsilon_{lij}(t) \\ m_{lij}(t) = \boldsymbol{x}_{lij}^T(t)\boldsymbol{\beta} + b_l + \boldsymbol{z}_{lij}^T(t)\boldsymbol{U}_{li} \end{cases} \tag{3.1}$$

where $m_{lij}(t)$ corresponds to the individual linear predictor of observed outcome. $\boldsymbol{x}_{lij}^T(t)$ and $\boldsymbol{z}_{lij}^T(t)$ are row-vectors of covariates with associated fixed effect coefficient $\boldsymbol{\beta}$ and random effects $\boldsymbol{U}_{li}$, respectively. The random intercept b_l is set for centers by assuming $b_l \sim N(0, \sigma_b^2)$.

The measurement error $\boldsymbol{\epsilon}_{li}$ is distributed as $N(\mathbf{0}, \boldsymbol{R}_{li})$. As proposed in Laird and Ware (1982), the $\boldsymbol{U}_{li}$ are distributed as $N(\mathbf{0}, \boldsymbol{D})$, allowing for $\boldsymbol{D}$ as an $n_{li} \times n_{li}$ positive-definite covariance matrix. Further simplification arises when $\boldsymbol{R}_{li} = \sigma^2 \boldsymbol{I}$, where $\boldsymbol{I}$ denotes an identity matrix and $\boldsymbol{D} = \boldsymbol{\Sigma_u}$ with the variance terms $\sigma^2_{u0}, \sigma^2_{u1}$ and the correlation coefficient term ρ. In addition, we assume $\boldsymbol{b}, \boldsymbol{U}, \boldsymbol{\epsilon}$ are mutually independent.

3.2.2 Time-to-recurrent Event Submodel

Let $T_{li} = min(T^*_{li}, C_{li})$ denote a terminal event time where T^*_{li} is the so-called 'true' event time and C_{li} is the corresponding censoring time. We define t_{lik} as the observed time for the event submodel with $t_{lik} \leq T_{li}$ $(k \in j)$ and d_{lik} as the event indicator. The parametric proportional hazard regression model is given by,

$$h_{lik}(t) = h_{l0}(t) exp\big\{\boldsymbol{\omega}^T_{lik}(t)\boldsymbol{\gamma} + f(\boldsymbol{\beta}, b_l, \boldsymbol{U}_{li}, \alpha_l; t) + v_{li}\big\} \tag{3.2}$$

where $h_{lik}(t)$ denotes $h_{li}(t_{lik})$ for simplicity, $h_{l0}(t)$ is the center-specific baseline hazard, $\boldsymbol{\omega}^T_{lik}(t)$ is a row-vector of individual covariates (possibly time-dependent) with corresponding regression coefficient $\boldsymbol{\gamma}$ (also known as log hazard ratio). The longitudinal and event processes are assumed to be related via an 'association structure', which is denoted by $f(\cdot)$. In particular, center-specific α_l quantifies the association between the time-varying longitudinal marker and the risk of a recurrent event. Based on our preliminary analysis, we choose current value and current slope of the linear predictor as the association structure, such as

$$\begin{cases} \text{Current value: } f(\boldsymbol{\beta}, b_l, \boldsymbol{U}_{li}, \alpha_{v_l}; t) = \alpha_{v_l} \times m_{lik}(t) \\ \text{Current slope: } f(\boldsymbol{\beta}, b_l, \boldsymbol{U}_{li}, \alpha_{s_l}; t) = \alpha_{s_l} \times \frac{d}{dt} m_{lik}(t) \end{cases} \tag{3.3}$$

The term v_{li} is a random effect that accounts for the correlation between recurrent events, which is assumed to follow $v_{li} \sim N(0, \sigma_v^2)$ independently. Note that $exp(v_{li})$ is well known as the frailty term in the survival context. We further assume that our longitudinal data is non-informative right censoring and the baseline hazard function is modeled with a Weibull distribution, that is $h_{l0}(t) = \delta_l t^{\delta_l - 1}$ with δ_l as center-specific shape parameter.

3.2.3 Conditional Independence

Let $\boldsymbol{\theta}$ denote a vector of all unknown parameters, we have following expressions by the assumption of conditional independence:

The longitudinal process is conditionally independent of the event process,

$$p(\boldsymbol{t}_{li}, \boldsymbol{d}_{li}, \boldsymbol{y}_{li} | b_l, \boldsymbol{U}_{li}, v_{li}; \boldsymbol{\theta}) = p(\boldsymbol{t}_{li}, \boldsymbol{d}_{li} | b_l, \boldsymbol{U}_{li}, v_{li}; \boldsymbol{\theta}) \times p(\boldsymbol{y}_{li} | b_l, \boldsymbol{U}_{li}; \boldsymbol{\theta}) \tag{3.4}$$

Repeated measurements in the longitudinal process are independent of each other,

$$p(\boldsymbol{y}_{li} | b_l, \boldsymbol{U}_{li}; \boldsymbol{\theta}) = \prod_j p(y_{lij} | b_l, \boldsymbol{U}_{li}; \boldsymbol{\theta}) \tag{3.5}$$

Recurrent events in the event process are independent of each other,

$$p(\boldsymbol{t}_{li}, \boldsymbol{d}_{li} | b_l, \boldsymbol{U}_{li}, v_{li}; \boldsymbol{\theta}) = \prod_k p(t_{lik}, d_{lik} | b_l, \boldsymbol{U}_{li}, v_{li}; \boldsymbol{\theta}) \tag{3.6}$$

3.2.4 Bayesian Inference

In this section, we specify the posterior distribution, likelihood function of event submodel and prior distributions. All posterior samplings are carried out from the new lightweight R interface CmdStanr to Stan (Gelman et al. (2013a), Stan Development Team (2011-2019)). In both simulation and application study, we obtain 2000 post-warmup posterior samplings via two chains and ensure the convergence by Gelman and Rubin potential scale reduction statistic $\widehat{R}$. Details of elapsed time with system information are included in Appendix C.6.

3.2.4.1 Posterior distribution

Under the full conditional independence assumption, the individual posterior distribution can be specified as,

$$p(\boldsymbol{\theta}, b_l, \boldsymbol{U}_{li}, v_{li}|\boldsymbol{y}_{li}, \boldsymbol{t}_{li}, \boldsymbol{d}_{li}) \propto \Big[\prod_j p(y_{lij}|b_l, \boldsymbol{U}_{li}, \boldsymbol{\theta}) \prod_k p(t_{lik}, d_{lik}|b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta})\Big] p(b_l|\boldsymbol{\theta}) p(\boldsymbol{U}_{li}|\boldsymbol{\theta}) p(v_{li}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \tag{3.7}$$

Specifically, the log likelihood function for the event submodel under two risk scenarios can be rewritten as

$$\begin{cases} \text{Calendar time: } \log p(t_{lik}, d_{lik}|b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta}) = d_{lik} \cdot \log h_{li}(t_{lik}) - \int_{t_{li(k-1)}}^{t_{lik}} h_{li}(s)ds \\ \text{Gap time: } \log p(t_{lik}, d_{lik}|b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta}) = d_{lik} \cdot \log h_{li}(t_{lik} - t_{li(k-1)}) - \int_0^{t_{lik}-t_{li(k-1)}} h_{li}(s)ds \end{cases} \tag{3.8}$$

where the latter integral term can be evaluated approximately by Gauss-Kronrod quadrature with Q nodes (Laurie (1997)), such that

$$\int_a^b h(s)ds \approx \sum_{q=1}^{Q} w_{q,scaled} \cdot h(s_{q,scaled}) \tag{3.9}$$

where $w_{q,scaled} = w_q \cdot \frac{b-a}{2}$ and $s_{q,scaled} = \frac{s_q+1}{2} \cdot (b-a) + a$ are scaled weights and locations for quadrature node $q(q = 1, \ldots, Q)$. The w_q and s_q are standardised weights and locations (also known as abscissa) on interval $[-1, 1]$ and we obtain their specific values from the source code of stan_jm from R package rstanarm by choosing quadrature nodes $Q = 7$.

3.2.4.2 Prior specifications

We identify mutually independent diffuse or weakly informative but proper priors for all unknown parameters in the proposed joint model. Let $N(\mu, \sigma^2)$ denote normal distribution with location $\mu \in \mathbb{R}$ and scale $\sigma \in \mathbb{R}^+$; $t(\nu, \mu, \sigma)$ denote student's t distribution with degrees of freedom $\nu \in \mathbb{R}^+$, location $\mu \in \mathbb{R}$ and scale $\sigma \in \mathbb{R}^+$; $lkjCorr(\Sigma|\eta) \propto det(\Sigma)^{\eta-1}$ represent Lewandowski-Kurowicka-Joe (LKJ) correlation distribution, such that Σ is a positive-definite, symmetric matrix with unit diagonal correlation matrix (i.e., a correlation matrix) with shape parameter $\eta \in \mathbb{R}^+$ (see Lewandowski et al. (2009) for details). The prior specifications for the simulation and the application studies are:

$$\beta_0 \sim N(0, 100^2),$$

$$\beta_p \sim N(0, \phi_\beta^2), p = 1, 2, \ldots, P,$$

$$\sigma \sim N(0, \phi_\sigma^2),$$

$$\sigma_b \sim N(0, 10^2),$$

$$\sigma_{u0}, \sigma_{u1} \sim t(1, 0, 10),$$

$$\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right) \sim lkjCorr(2),$$

$$\gamma_0 \sim N(0, 20^2),$$

$$\gamma_q \sim N(0, \phi_\gamma^2), q = 1, 2, \ldots, Q,$$

$$\lambda_l \sim N(0, 5^2), l = 1, 2, \ldots, L,$$

$$\sigma_v \sim N(0, 10^2)$$

where ϕ_β and ϕ_γ are standard deviations of corresponding design matrices and ϕ_σ is the standard deviation of observed longitudinal outcomes. Practically, Stan provides an implicit parameterization of the LKJ correlation matrix density in terms of its Cholesky factor, thus we can set L_u as a Cholesky factor of the correlation matrix $\Sigma = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$, such that $L_u \sim$ lkj_corr_cholesky(2) implies $\Sigma = L_u \cdot L_u^T \sim lkjCorr(2)$. Readers who are interested in this topic can refer to Sorensen et al. (2016) for generating correlated random variables using the Cholesky decomposition.

3.2.5 Individual Prediction

The individual prediction for the longitudinal marker at time t, can be generated from the posterior predictive distribution

$$p(\tilde{y}_{lij}(t)|\mathcal{D}) = \int\int\int p(\tilde{y}_{lij}(t)|b_l, \boldsymbol{U}_{li}, \boldsymbol{\theta}) p(b_l, \boldsymbol{U}_{li}, \boldsymbol{\theta}|\mathcal{D})\, db_l\, d\boldsymbol{U}_{li}\, d\boldsymbol{\theta} \tag{3.10}$$

where $\mathcal{D} = \{\boldsymbol{y}_{li}, \boldsymbol{t}_{li}, \boldsymbol{d}_{li}; l = 1, \ldots, L, i = 1, \ldots, n_l\}$ is the entire collection of data. To compute Equation 3.10, we can draw samplings from $p(y_{lij}(t)|b_l^{(m)}, \boldsymbol{U}_{li}^{(m)}, \boldsymbol{\theta}^{(m)})$, such that $b_l^{(m)}, \boldsymbol{U}_{li}^{(m)}$ and $\boldsymbol{\theta}^{(m)}$ are the $m^{th}(m = 1, \ldots, M)$ HMC draws from the joint posterior distribution of $p(b_l, \boldsymbol{U}_{li}, \boldsymbol{\theta}|\mathcal{D})$ as

$$p(\tilde{y}_{lij}(t)|\mathcal{D}) \approx \frac{1}{M}\sum_{m=1}^{M} p(\tilde{y}_{lij}(t)|b_l^{(m)}, \boldsymbol{U}_{li}^{(m)}, \boldsymbol{\theta}^{(m)}, \mathcal{D}). \tag{3.11}$$

In parallel, for i^{th} individual who has $n_{li}^*(n_{li}^* = 0, 1, 2, \ldots)$ recurrent events up to the time t, it might be of interest to look beyond and predict the probability of next event-free outcome in the time frame $(t, t']$ with $t' = t + \Delta t$. For this cause, the conditional survival (PEx-free) probability can be written as,

$$\begin{aligned} S_{li}(t'|t) &= p(t_{n_{li}^*+1} \geq t'|t_{n_{li}^*+1} > t, \mathcal{D}) \\ &= \int\int\int\int p(t_{n_{li}^*+1} \geq t'|t_{n_{li}^*+1} > t, b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta}, \mathcal{D}) \\ &\quad \cdot p(b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta}|t_{n_{li}^*+1} > t, \mathcal{D})\, db_l\, d\boldsymbol{U}_{li}\, dv_{li}\, d\boldsymbol{\theta} \\ &\approx \frac{1}{M}\sum_{m=1}^{M} exp\Big[-\int_t^{t'} h\big(s|b_l^{(m)}, \boldsymbol{U}_{li}^{(m)}, v_{li}^{(m)}, \boldsymbol{\theta}^{(m)}\big) ds\Big] \end{aligned} \tag{3.12}$$

where the integration respect to $\{b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta}\}$ is approximated using Monte Carlo method from their posterior samples. The comprehensive derivation of Equation 3.12 can be found in

Appendix C.1. The integral term of $\int_t^{t'} h(s|\cdot)ds$ can be approximated by Gauss-Kronrod with Q=15 quadrature nodes. The 2.5% and 97.5% quantiles of posterior draws from Equation 3.12 can be obtained as the credible intervals for the predictive probability.

3.2.6 Predictive Performance

To assess the predictive performance, we employ area under receiver operating characteristic curve (hereafter, AUC) for discrimination (discriminate between individuals who will experience the next recurrent event from subjects who will not) and mean predictive error (MPE) for calibration (how well the model predicts the observed event probability). We calculate time-dependent AUC and MPE (e.g., under squared loss function) based on source code predictive_accuracy.stanjm from R package rstanarm and some other nice references (Andrinopoulou et al. (2018), Andrinopoulou et al. (2021)). The details are described in Table 3.1.

Table 3.1: Algorithm for time-dependent AUC and MPE

Time-dependent AUC

1. Define individual-specific start time $t_i = \text{tstart}_i$ and a common future stop time t'. To conform the prediction data, we only include individuals who are still at risk of the event at t. For longitudinal data, we adopt observations observed until t_i.
2. Calculate event-free ('survival') probability at t' and observed tstop_i for each individual based on Equation 3.12 to obtain $S_i(t'|t_i)$ and $S_i(\text{tstop}_i|t_i)$
3. Sort individuals by their observed tstop in an increasing order and group each two by combinations without replacement.
4. AUC is calculated by accounting for weights caused by censoring conditions. For each combation $c = 1, \ldots, C$, assume that $\text{tstop}_i < \text{tstop}_j$:
   - If only individual i is censored at t', which means $\text{tstop}_i \leq t'$ & $\text{status}_i = 0$, then weight $w = 1 - S_i(\text{tstop}_i|t_i)$
   - If only individual j is censored at t', which means $\text{tstop}_j \leq t'$ & $\text{status}_j = 0$, then weight $w = S_j(\text{tstop}_j|t_j)$
   - If both individuals are censored at t', then weight $w = \big(1 - S_i(\text{tstop}_i|t_i)\big) \times S_j(\text{tstop}_j|t_j)$
   - If it does not belong to above cases, $w = 1$
   - Let $S_i = S_i(t'|t_i), S_j = S_j(t'|t_j)$, compute $A_c = I_{S_i<S_j} \cdot w$; $D_c = I_{S_i>S_j} \cdot w$; $T_c = I_{S_i=S_j} \cdot w$, where I_x denotes a indicator function with 1 when x is true and 0, otherwise.
5. Repeat Step 4 until C times
6. $$\text{AUC} = \sum_{c=1}^{C} \Big(\frac{A_c + 0.5 \cdot T_c}{A_c + D_c + T_c} \Big)$$

Time-dependent MPE

1. For each individual $i (i = 1, \ldots, N)$:
   - If individual i died or censored after t', $\text{Error}_i = (1 - S_i)^2$
   - If individual i died before t', $\text{Error}_i = (0 - S_i)^2$
   - If individual i censored before t', $\text{Error}_i = S_i(\text{tstop}_i|t_i) \times (1 - S_i)^2 + \big(1 - S_i(\text{tstop}_i|t_i)\big) \times (0 - S_i)^2$
2. $$\text{MPE} = \sum_{i=1}^{N} \text{Error}_i / N$$

3.2.7 Model Selection

We allude to the leave-one-out (LOO) cross-validation for model choice, which in its own words is the method for estimating pointwise out-of-sample prediction accuracy from the log-likelihood evaluated at the posterior simulations of the parameter values. Vehtari et al. (2017) developed an efficient computation for LOO using pareto-smoothed importance sampling (PSIS) to lay out fast and stable computations for LOO. PSIS-LOO is demonstrated to be even more robust than asymptotic widely applicable information criterion (WAIC) (Watanabe (2010)), which is a recent popular measure of predictive accuracy. We implement the computations of LOO information criterion (LOOIC) via R package loo (v2.3.1, Vehtari et al. (2020)) by extracting log likelihoods from posterior samplings.

The PSIS estimate of expected log pointwise predictive density (elpd) is

$$\widehat{\text{elpd}}_{\text{psis-loo}} = \sum_{l=1}^{L} \sum_{i=1}^{n_l} log\Big(\frac{\sum_{s=1}^{S} w_{li}^{s} p(y_{li}|\theta^s)}{\sum_{s=1}^{S} w_{li}^{s}}\Big) \tag{3.13}$$

where w_{li}^{s} denotes weights at iteration s, $p(y_{li}|\theta^s)$ denotes likelihood function at s. As with WAIC, we define LOOIC in the Equation 3.14 so as to be on the deviance scale. The model with smaller LOOIC value indicates the better goodness of fit.

$$\text{LOOIC} = -2 \times \widehat{\text{elpd}}_{\text{psis-loo}} \tag{3.14}$$

3.3 Simulation Study

We simulate a total of four hierarchical data sets, with the aim to assess the performance of four proposed joint model in a comparison of corresponding two-stage method. These proposed joint models are distinguished by the association structure under the risk scale as illustrated in Table 3.2. For the two-stage method, in stage one, longitudinal ppFEV1 are modeled by LME model, which, in stage two, the individual predicted 'unobserved and true' ppFEV1 is used as a covariate in the time-to-recurrent event model.

Table 3.2: Simulation illustration

Association + Risk scale	**Simulated Data**	**Fitted Model**
Slope + Gap	JM1	JM1 TM1
Slope + Calendar	JM2	JM2 TM2
Value + Gap	JM3	JM3 TM3
Value + Calendar	JM4	JM4 TM4

Note:JM=Joint Model; TM=Two-stage Method

We simulate each data set consisting of an average of 480 individuals who are from $L = 6$ centers for 50 replicates for each proposed joint model. The data generating algorithm is summarized in Table 3.3. It is worth noting that we add a fixed time window (28 days) to mimic the real case by accounting for patient's recovery time. Comparisons between estimate and true value of each parameter are displayed in Figure 3.3 and Figure 3.2. The joint models represent excellent performance in recovering the true values, which validate

the Bayesian inference, while two-stage approaches yield to slight biases. Details for explicit estimates are described in Appendix C.2.

Table 3.3: Simulation algorithm

Random effects

Simulate $b_l \sim N(0, \sigma_b^2)$, $\boldsymbol{U}_{li} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_u)$, and $\boldsymbol{\Sigma}_u$ is the covariance matrix with the variance terms $\sigma_{u0}^2, \sigma_{u1}^2$ and the correlation coefficient term ρ

Event Process

1. Terminal event
   - Simulate $\omega_1 \sim \text{Bernoulli}(0.5), \omega_2 \sim N(0,1), v_{li} \sim N(0, \sigma_v^2), u_{li} \sim \text{Uniform}(0,1)$
   - Let $A = \gamma_0 + \gamma_1\omega_1 + \gamma_2\omega_2$
   - Define hazard function with baseline Weibull hazard $h_{li}(t) = \delta_l t^{\delta_l - 1} exp(A)$
   - Cumulative hazard function $H_{li}(t) = \int_0^t h_{li}(s)ds$
   - Solve t from equation $H_{li}(t) + log(u_{li}) = 0$
   - Terminal time $T_{li} = min(t, t_{\max})$, where maximum follow-up time $t_{\max} = 10$ years
2. Recurrent event
   - Let $f(\beta, b_l, \boldsymbol{U}_{li}; t) = \alpha_{vl} \cdot m_{ik}(t)$ or $\alpha_{sl} \cdot m'_{ik}(t)$, where $m_{ik}(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + b_l + U_{0li} + U_{1li} t$
   - Hazard function of k^{th} event: $h_{lik}(t) = \delta_l t^{\delta_l - 1} exp[A + f(\beta, b_l, \boldsymbol{U}_{li}; t) + v_{li}]$
   - Set starting point: $\text{tstart}_{li0} = 0$
   - In each loop $k = 1, \ldots, K$, generate $u_{lik} \sim \text{Uniform}(0,1)$
   - If gap time: Solve Δt_{lik} from equation $H_{li}(\Delta t_{lik}) + log(u_{lik}) = 0$, where $H_{li}(\Delta t_{lik}) = \int_0^{\Delta t_{lik}} \delta_l s^{\delta_l - 1} exp\big[A + f(\beta, b_l, \boldsymbol{U}_{li}; s + \text{tstart}_{lik}) + v_{li}\big] ds$
   - If calendar time: Solve Δt_{lik} from equation $H_{li}(\Delta t_{lik}) + log(u_{lik}) = 0$, where $H_{li}(\Delta t_{lik}) = \int_0^{\Delta t_{lik}} \delta_l (s + \text{tstart}_{lik})^{\delta_l - 1} exp\big[A + f(\beta, b_l, \boldsymbol{U}_{li}; s + \text{tstart}_{lik}) + v_{li}\big] ds$
   - Update $\text{tstart}_{lik} = \text{tstart}_{li(k-1)} + \Delta t_{lik} + 28$ days, $\text{tstop}_{li(k-1)} = \text{tstart}_{li(k-1)} + \Delta t_{lik}$ until tstart_{lik} reaches up to T_{li}

Longitudinal Process

Once $\text{tstop}_{li(k-1)}$ is obtained from the previous step, we adjust $t_{lij} = \text{tstop}_{li(k-1)}$ for $j = 2, \ldots, K+1$ and set $t_{lij} = 0$ for $j = 1$, such that $y_{lij} = m_{lij}(t_{lij}) + \epsilon_{lij}$, where $\epsilon_{lij} \sim N(0, \sigma^2)$

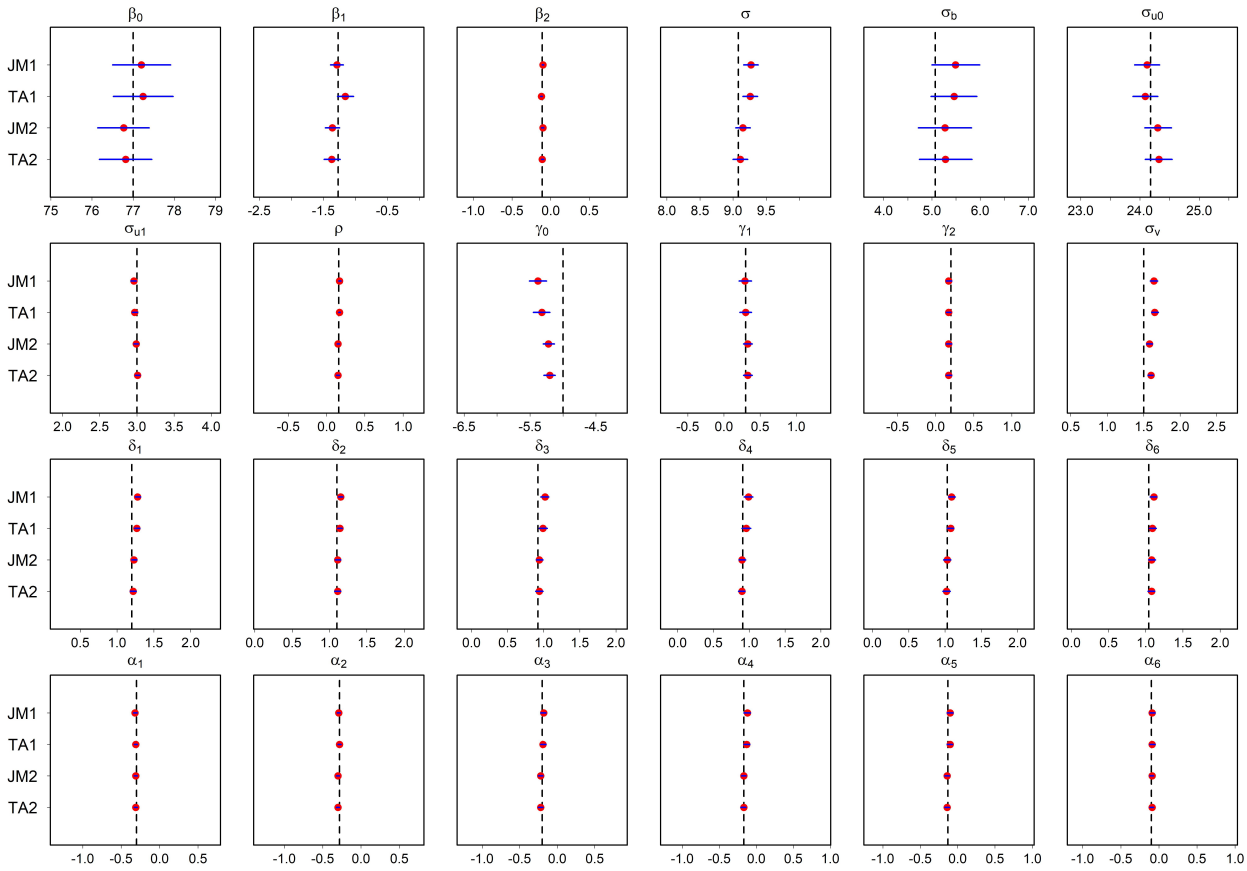

Figure 3.2: Simulation results based on 50 replicates with true value (dashed vertical line), posterior mean estimate (red dot) and corresponding 95% confidence interval (blue line). JM=Joint Model; TM=Two-stage Method; JM3/TM3: Value+Gap; JM4/TM4: Value+Calendar

3.4 Application of CF Study

In this section, we illustrate our motivating multi-center CF study, which tracked the lung trajectory and corresponding covariates of patients since 1997 by US CFFPR and this study was approved by the IRB at CCHMC and UC.

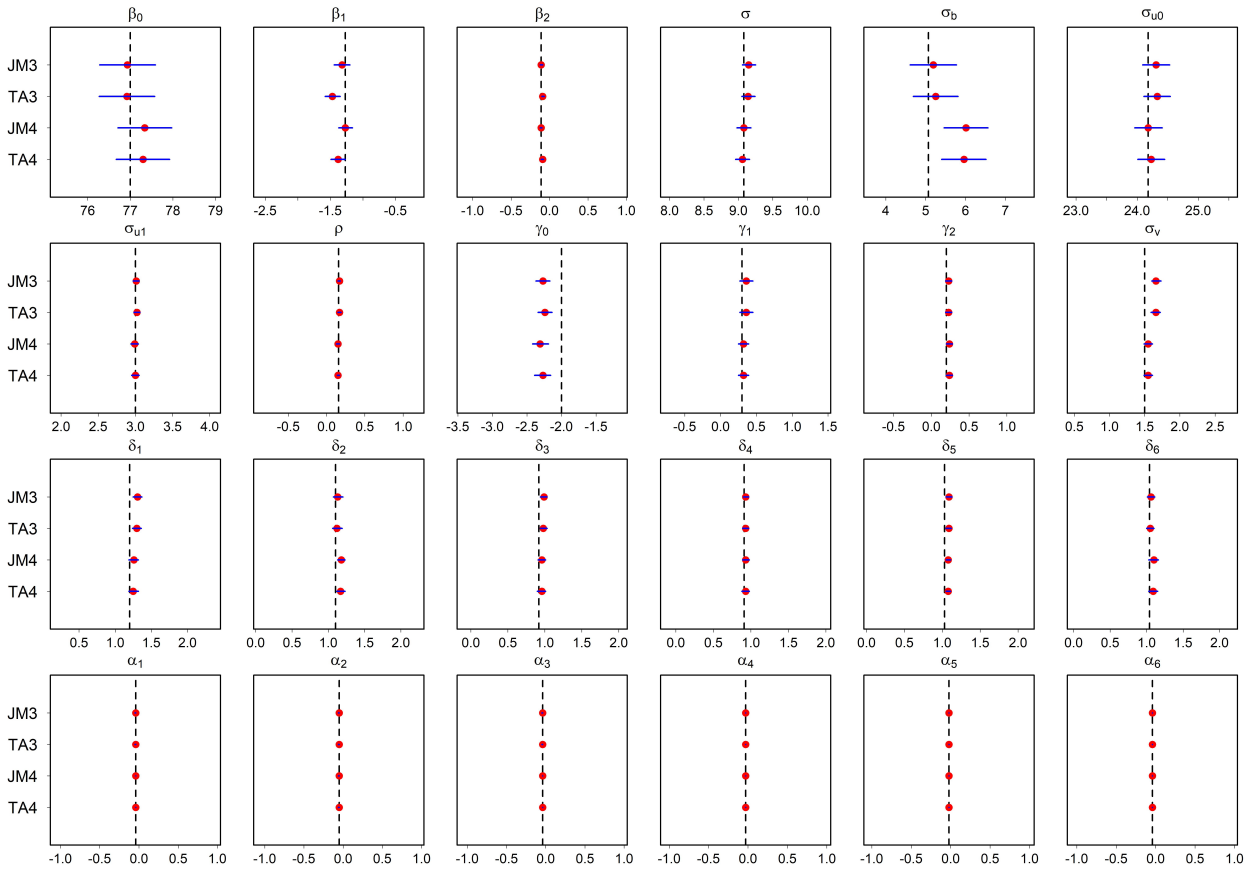

Figure 3.3: Simulation results based on 50 replicates with true value (dashed vertical line), posterior mean estimate (red dot) and corresponding 95% confidence interval (blue line). JM=Joint Model; TM=Two-stage Method; JM1/TM1: Slope+Gap; JM2/TM2: Slope+Calendar

3.4.1 Motivating Data

The details of data cleaning process are described in Appendix C.3. Unlike five random centers selected (which consist of 381 patients) in Chapter 2, we categorize all centers into three levels (e.g., mild, moderate, severe), in which we randomly choose two centers, with the aim to characterize heterogeneity. Therefore, the final cleaned data we obtained is of 440 patients from 6 centers, which 178 (40.45%) patients never encounter PEx, 7 (1.59%) patients experi-

ence PEx only once, 255 (57.95%) patients experience recurrent events during the follow-up period. Descriptive statistics by centers are summarized in Table 3.4, of which includes: i) demographic summary, such as Gender, Genotype, Birth cohort; ii) the first recorded covariates, such as baseline age and baseline ppFEV1; iii) binary time-varying measures on insurance use, microbiology (infection with pa or MRSA) and CF-related diabetes mellitus (cfrd). The time since baseline is utilized as the time scale for analysis. The category levels of birth cohort are computed by quartile 0-25%, 25%-50%, 50%-75% and 75%-100%, respectively. This covariate may offer some remedies for biases induced by left-truncation from review year 2003. To account for irregular visit due to disease severity, we have included numbers of PEx and outpatient visits within the year prior to a given clinical encounter as predictors. Trajectory in Figure 3.4 displays the heterogeneous nature of lung function and recurrent feature of PEx over the life span. We note that PEx events are more likely to appear when the averaged values of ppFEV1 are low.

Table 3.4: Demographic clinical summary across centers

	Center 1 (N=74)	Center 2 (N=70)	Center 3 (N=75)	Center 4 (N=75)	Center 5 (N=72)	Center 6 (N=74)
Baseline age (years)						
Mean; Median	19.0; 17.0	29.4; 21.9	15.7; 13.5	17.7; 14.7	11.3; 10.6	17.4; 15.0
(Min - Max)	(6.08 - 47.6)	(17.7 - 79.8)	(6.11 - 63.9)	(6.04 - 53.4)	(6.02 - 34.6)	(6.01 - 51.9)
Baseline ppFEV1						
Mean; Median	77.4; 79.2	64.7; 65.8	80.5; 83.5	82.2; 88.7	80.4; 85.8	79.6; 80.4
(Min - Max)	(16.2 - 144)	(25.3 - 111)	(23.7 - 134)	(19.7 - 117)	(18.3 - 129)	(19.6 - 129)
Gender						
Female	23 (31.1%)	31 (44.3%)	31 (41.3%)	30 (40.0%)	24 (33.3%)	25 (33.8%)
Male	51 (68.9%)	39 (55.7%)	44 (58.7%)	45 (60.0%)	48 (66.7%)	49 (66.2%)
GenotyPEx (F508del)						
Neither/Unknown	30 (40.5%)	14 (20.0%)	10 (13.3%)	17 (22.7%)	14 (19.4%)	13 (17.6%)
Homozygous	11 (14.9%)	22 (31.4%)	40 (53.3%)	28 (37.3%)	30 (41.7%)	28 (37.8%)
Heterozygous	33 (44.6%)	34 (48.6%)	25 (33.3%)	30 (40.0%)	28 (38.9%)	33 (44.6%)
Birth cohort						
< 1988	16 (21.6%)	26 (37.1%)	20 (26.7%)	28 (37.3%)	5 (6.9%)	30 (40.5%)
$[1988, 1993)$	14 (18.9%)	21 (30.0%)	13 (17.3%)	19 (25.3%)	20 (27.8%)	13 (17.6%)
$[1993, 1998)$	12 (16.2%)	23 (32.9%)	13 (17.3%)	11 (14.7%)	19 (26.4%)	16 (21.6%)
> 1998	32 (43.2%)	0 (0%)	29 (38.7%)	17 (22.7%)	28 (38.9%)	15 (20.3%)
Insurance use						
At baseline	44 (59.5%)	8 (11.4%)	49 (65.3%)	58 (77.3%)	33 (45.8%)	37 (50.0%)
Ever during follow-up	53 (71.6%)	17 (24.3%)	70 (93.3%)	68 (90.7%)	48 (66.7%)	50 (67.6%)
Pseudomonas aeruginosa (pa)						
Baseline	19 (25.7%)	15 (21.4%)	18 (24.0%)	22 (29.3%)	15 (20.8%)	17 (23.0%)
Ever follow-up	36 (48.6%)	41 (58.6%)	49 (65.3%)	47 (62.7%)	54 (75.0%)	47 (63.5%)
Methicillin-resistant Staphylococcus aureus (MRSA)						
At baseline	18 (24.3%)	2 (2.9%)	11 (14.7%)	5 (6.7%)	7 (9.7%)	3 (4.1%)
Ever during follow-up	28 (37.8%)	19 (27.1%)	41 (54.7%)	21 (28.0%)	37 (51.4%)	30 (40.5%)
CF-related diabetes mellitus (cfrd)						
At baseline	9 (12.2%)	16 (22.9%)	6 (8.0%)	8 (10.7%)	7 (9.7%)	7 (9.5%)
Ever during follow-up	22 (29.7%)	24 (34.3%)	32 (42.7%)	33 (44.0%)	23 (31.9%)	33 (44.6%)
On Enzymes						
At baseline	53 (71.6%)	58 (82.9%)	38 (50.7%)	27 (36.0%)	30 (41.7%)	25 (33.8%)
Ever during follow-up	60 (81.1%)	61 (87.1%)	68 (90.7%)	69 (92.0%)	71 (98.6%)	63 (85.1%)
PEx event						
At baseline	22 (29.7%)	25 (35.7%)	14 (18.7%)	17 (22.7%)	29 (40.3%)	17 (23.0%)
Ever during follow-up	49 (66.2%)	30 (42.9%)	42 (56.0%)	42 (56.0%)	58 (80.6%)	39 (52.7%)

Note: N=num patient

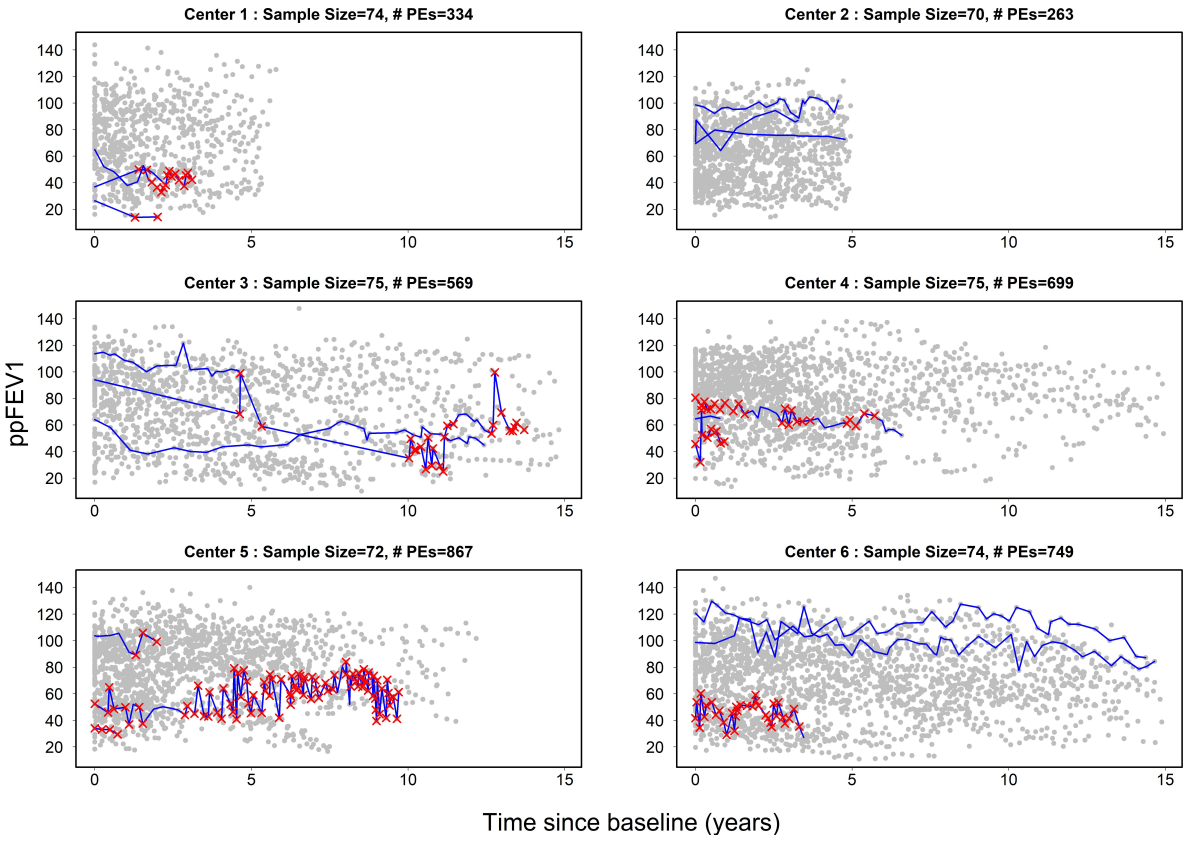

Figure 3.4: Observed ppFEV1 against time since baseline (in years) for each center. Within each center: three random profiles (blue lines), observed values (gray dots) and recurrent PEx events (red crosses)

3.4.2 Parameter Estimation

For the sake of internal validation, we split our data into training and masking cohorts without loss of patients. For patients who do not experience PEx event, we mask last period of their observations in terms of quantiles, while for rest of patients, we mask their last observed PEx event. Analogously to the Chapter 2, we select predictors from aforementioned variables by Stepwise method by a two-stage approach. To the end, we consider the following risk factors: i) For longitudinal submodel: baseline ppFEV1, time, quadratic time, number

of PEx within prior year, birth cohort, infection with pa and gender; ii) For event submodel: number of previous PEx, insurance use at baseline and gender. We examine a total of eight models (four JMs and four TMs) and their comparisons with respect to LOOIC are represented in Table 3.5. Undoubtedly, all joint models are superior than corresponding models by two-stage approach. From the statistical sight, the joint model with current value of ppFEV1 as the association structure in a calendar time scale outperforms others with the lowest LOOIC. Nonetheless, the choice between calendar time and gap time also depends on clinical interests and predictive performance. For illustration purpose, we pick the 'optimal' model from statistical perspective and its estimations with 90% equal-tail credible intervals (CI) are summarized in Table 3.6 and convergence diagnosis plots are included in Appendix C.5.

From results of the longitudinal submodel, having higher ppFEV1 at entry corresponds to a higher ppFEV1 against time which reconciles with previous findings (Taylor-Robinson et al. (2012), Li et al. (2017)). The presence of infections with pa corresponds to worsen an average -1.15 (90% CI [-1.64, -0.66]) units of ppFEV1. Being a male patient in older birth cohort is associated with higher ppFEV1. The variance component estimates indicate substantial heterogeneity within and between patients, in terms of large $\sigma, \sigma_{u0}, \sigma_{u1}$. While less variations are found between centers given a frail σ_b. For the event submodel, we observe that males had 28.11% $(exp(\gamma_3) - 1)$ lower risk to encounter PEx compared to females. The usage of insurance at entry increased the risk by 33.64% $(exp(\gamma_2) - 1)$, though it might not be statistically significant. One more count added to cumulative PEx events is associated

with a 1.01% $(exp(\gamma_1) - 1)$ increase in the risk of current PEx. Center-specific association parameters are all negative. Specifically, every one percentage predicted increase in 'true and unobserved' ppFEV1 would decrease 3.92%, 4.88%, 3.92%, 3.92%, 2.96%, 3.92% for the PEx risk from Center 1 to Center 6 (mild-moderate-severe), respectively. If we evaluate the second-best joint model 'Slope+Gap' (Appendix C.4), every one unit rate of increase in ppFEV1 would decrease 28.8%, 40.5%, 37.5%, 37.5%, 34.9%, 24.4% for the PEx risk from Center 1 to Center 6 (mild-moderate-severe), respectively. The non-zero center-specific Weibull shape parameters and frailty term are believed to facilitate the flexibility of the joint model for multi-center cohorts.

Table 3.5: Model comparisons with the boldface as the smallest LOOIC

Association + Risk scale	Model	LOOIC [a](SE[b])	$LOOIC_1$[c](SE[b])	$LOOIC_2$[d](SE[b])
Slope + Gap	Joint Model	52551.9 (383.1)	52490.3 (231.5)	61.6 (151.6)
	Two-stage Method	52578.8 (385.7)	52460.9 (232.3)	117.9 (153.4)
Slope + Calendar	Joint Model	52563.0 (389.4)	52473.1 (231.9)	89.9 (157.5)
	Two-stage Method	52575.5 (390.2)	52460.9 (232.3)	114.6 (157.9)
Value + Gap	Joint Model	52574.8 (383.4)	52491.1 (230.6)	83.7 (152.8)
	Two-stage Method	52602.4 (384.5)	52460.9 (232.3)	141.5 (152.2)
Value + Calendar	**Joint Model**	**52537.3 (386.2)**	52486.4 (231.1)	50.9 (155.1)
	Two-stage Method	52551.3 (388.3)	52460.9 (232.3)	90.4 (156)

[a] Standard error approximated as byproduct of loo package; [c] Longitudinal submodel; [d] Event submodel

Table 3.6: Model estimations under Joint Model: Value+Calendar

	mean [1]	sd [2]	q5 [3]	q95 [3]	rhat [4]	ess_bulk [5]	ess_tail [6]
Longitudinal submodel							
β_0 (intercept at age 6 years)	7.59	1.59	5.09	10.15	1.00	713.45	1017.42
β_1 (baseline ppFEV1)	0.86	0.02	0.83	0.89	1.00	553.14	1026.96
β_2 (time since baseline)	-1.12	0.22	-1.48	-0.76	1.00	507.77	850.71
β_3 (time since baseline^2)	-0.11	0.02	-0.14	-0.09	1.00	1049.14	1517.04
β_4 (number of PExs within prior year)	-0.33	0.05	-0.4	-0.25	1.00	1742.33	1373.07
β_5 (birth cohort*$[1988, 1993)$)	3.17	1.15	1.28	5	1.00	592.67	961.08
β_6 (birth cohort $[1993, 1998)$)	3.48	1.22	1.32	5.46	1.00	503.65	885.29
β_7 (birth cohort >1998)	4.57	1.24	2.49	6.69	1.00	474.94	805.76
β_8 (pa)	-1.15	0.3	-1.64	-0.66	1.00	3771.81	1534.25
β_9 (gender*male)	1.59	0.83	0.23	2.93	1.01	470.46	924.86
σ (measurement error)	9.07	0.08	8.94	9.2	1.00	2531.6	1556.46
σ_b (between centers, intercept)	1.65	1.02	0.37	3.59	1.00	396.71	382.91
σ_{u0} (between patients, intercept)	6.78	0.31	6.27	7.29	1.00	784.14	1288.57
σ_{u1} (between patients, slope)	3.02	0.18	2.73	3.33	1.01	340	797.25
ρ (correlation, intercept and slope)	0.18	0.07	0.06	0.3	1.01	193.5	482.79
Event submodel							
γ_0 (intercept at age 6 years)	2.02	0.28	1.57	2.5	1.00	246.96	484.67
γ_1 (number of previous PExs)	0.01	0	0	0.02	1.00	1991.24	1683.3
γ_2 (baseline insurance use)	0.29	0.19	-0.02	0.6	1.00	307.03	651.51
γ_3 (gender*male)	-0.33	0.19	-0.64	-0.02	1.00	265.64	564.24
δ_1(weibull shape, Center 1)	1.24	0.09	1.1	1.38	1.00	1975.36	1535.57
δ_2(weibull shape, Center 2)	1.01	0.08	0.88	1.13	1.00	2476.21	1642.13
δ_3(weibull shape, Center 3)	1.28	0.07	1.16	1.4	1.00	754.97	1167.62
δ_4(weibull shape, Center 4)	1.26	0.06	1.16	1.36	1.00	1780.2	1785.08
δ_5(weibull shape, Center 5)	1.24	0.06	1.15	1.34	1.00	1466.17	1421.63
δ_6(weibull shape, Center 6)	1.07	0.05	0.98	1.15	1.00	1152.13	929.47
σ_v(between patients, frailty)	1.6	0.1	1.44	1.77	1.00	527.68	762.6
Association Structure							
α_1 (submodel link, Center 1)	-0.04	0	-0.05	-0.03	1.01	455.98	786.05
α_2 (submodel link, Center 2)	-0.05	0.01	-0.06	-0.04	1.00	331.54	639.72
α_3 (submodel link, Center 3)	-0.04	0	-0.05	-0.04	1.00	455.63	797.91
α_4 (submodel link, Center 4)	-0.04	0	-0.04	-0.03	1.01	359.67	666.75
α_5 (submodel link, Center 5)	-0.03	0	-0.03	-0.02	1.01	323.03	590.15
α_6 (submodel link, Center 6)	-0.04	0	-0.05	-0.04	1.01	381.33	895.84

[1] posterior mean; [2] Monte Carlo standard deviation; [3] posterior quantiles at 5% and 95%;
[4] potential scale reduction factor (at convergence, rhat=1) [5] bulk effective sample size;
[6] tail effective sample size; * Reference: Birth cohort < 1988; Gender female

3.4.3 Model Diagnostics

In this section, we validate the assumptions of our joint model (Value+Calendar) through basic residual tools by following the methodology addressed in Rizopoulos (2012b). For the longitudinal submodel, we implement conditional standardized residuals to check the assumptions of homoscedasticity and normality (see Plot A in Figure 3.5). We observe that there is no systematic trend from the fitted loess curve, despite some heavy-tailed behaviors (see Plot B in Figure 3.5). For the event submodel, we investigate the martingale residuals based on the counting process. Subject-specific martingale residual at time t can be viewed as the difference between the observed number of events and the expected number of events by t. The fitted zero loess curve illustrates the reasonable performance of the event submodel, despite some outliers. Another diagnostic method, which is called Cox-Snell residuals (Cox and Snell (1968)), can be utilized to compare the observed distribution of cumulative hazard with the expected one unit exponential distribution. Cox-Snell residuals are reserved for the time-to-event case, hence here we estimate the subject-specific cumulative hazard for the first PEx event. Kaplan-Meier estimate (Kaplan and Meier (1958)) of the survival function of the censored Cox-Snell residuals is illustrated in Plot D in Figure 3.5. We observe acceptable discrepancies between the fit of Kaplan-Meier estimate and the expected asymptotic distribution. Apart from previous Gauss-Kronrod approach, here we evaluate the numerical integrand via integrate function from R package stats (v4.0.2, R Core Team (2020)).

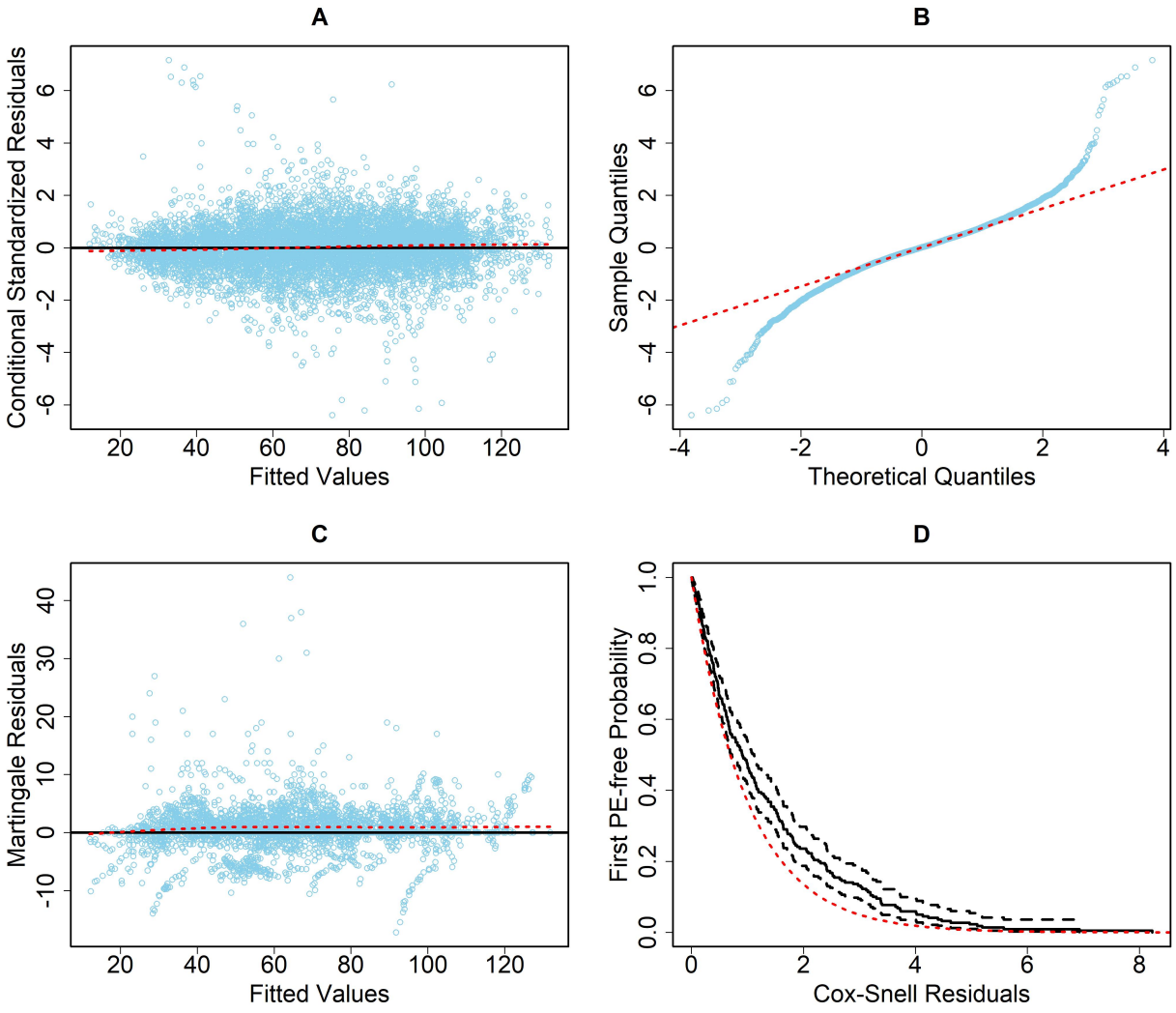

Figure 3.5: Residuals diagnostic plot for the proposed joint model. Upper panel: subject-specific standardized residuals versus fitted values (A) and normal Q-Q plot for longitudinal submodel (B). Lower panel: subject-specific martingale residuals versus fitted values (C) and Cox-Snell residuals for event submodel (D). Red dashed lines in A & C are fitted loess curve; Red dashed lines in B & D are normal curve and exponential curve, respectively

3.4.4 Predictive Performance

To assess predictive performance of proposed joint models, we compute time-dependent AUC and MPE for each joint model under various future time scenarios and results are shown in Table 3.7. Unlike the conventional method used in Ren et al. (2021), that is to assign a common prediction start time (t), we characterize individual prediction time (t_{li}) from the end time of each patient's last available encounter. The future time (t') is summarized by the quartile of their stop time. The concept of patients at risk means that patients have not yet experienced the next occurrence of PEx until t'.

All joint models represent the excellent discriminate and calibrate capability when more time-varying covariates become available in the longer term run. Figure 3.6 forecast random patients who are still at risk after their last observed measurements under the joint model Value+Calendar. We note that Patient 202 and Patient 284 represent two extreme cases, suggesting that higher lung function reduces the risk of the next PEx event, while the latter patient needs a timely care and prevention given high risk for the next PEx event. From virtual inspection, we observe that averaged ppFEV1 value and frequency of previous PEx occurrences are not ignorable risk factors for the prognostic probability of the next PEx event.

Table 3.7: Predictive performance of proposed joint models

t_{li}	Δt	t'	Num at risk	Joint Model	AUC	MPE
$[0, 2.73)$	1.41	2.73	42	Slope + Gap	0.61	0.28
				Slope + Calendar	0.64	0.27
				Value + Gap	0.66	0.27
				Value + Calendar	0.68	0.26
$[0, 5.10)$	1.56	5.10	65	Slope + Gap	0.89	0.14
				Slope + Calendar	0.88	0.14
				Value + Gap	0.90	0.14
				Value + Calendar	0.88	0.14
$[0, 7.84)$	2.76	7.84	31	Slope + Gap	0.92	0.12
				Slope + Calendar	0.91	0.13
				Value + Gap	0.92	0.12
				Value + Calendar	0.92	0.13

Note: t=individual prediction start time in year; Δt: averaged prediction window in year; t': future time in year; Num at risk: Number of patients at risk at t'; AUC=area under curve; MPE=mean predictive error based on squared loss function

3.5 Discussion

In this chapter, we have proposed a multilevel Bayesian joint model for a monitoring CF data depicted by hierarchical structures and irregularly clinical visits. Our novel model allows for center-specific association parameters to quantify the strength of the correlation between the two processes and such approach facilitate individual prognosis towards personalized medical monitoring. Furthermore, we provide detailed simulation algorithm and Stan codes by extending the work of Brilleman et al. (2018). Despite the high cost of computing time due to large sample size and approximation using Gauss-Kronrod quadrature rule, Bayesian methodology provides convincing benefits, especially lead to posterior predictive distributions and release the numerical integral burden through Monte Carlo samplings.

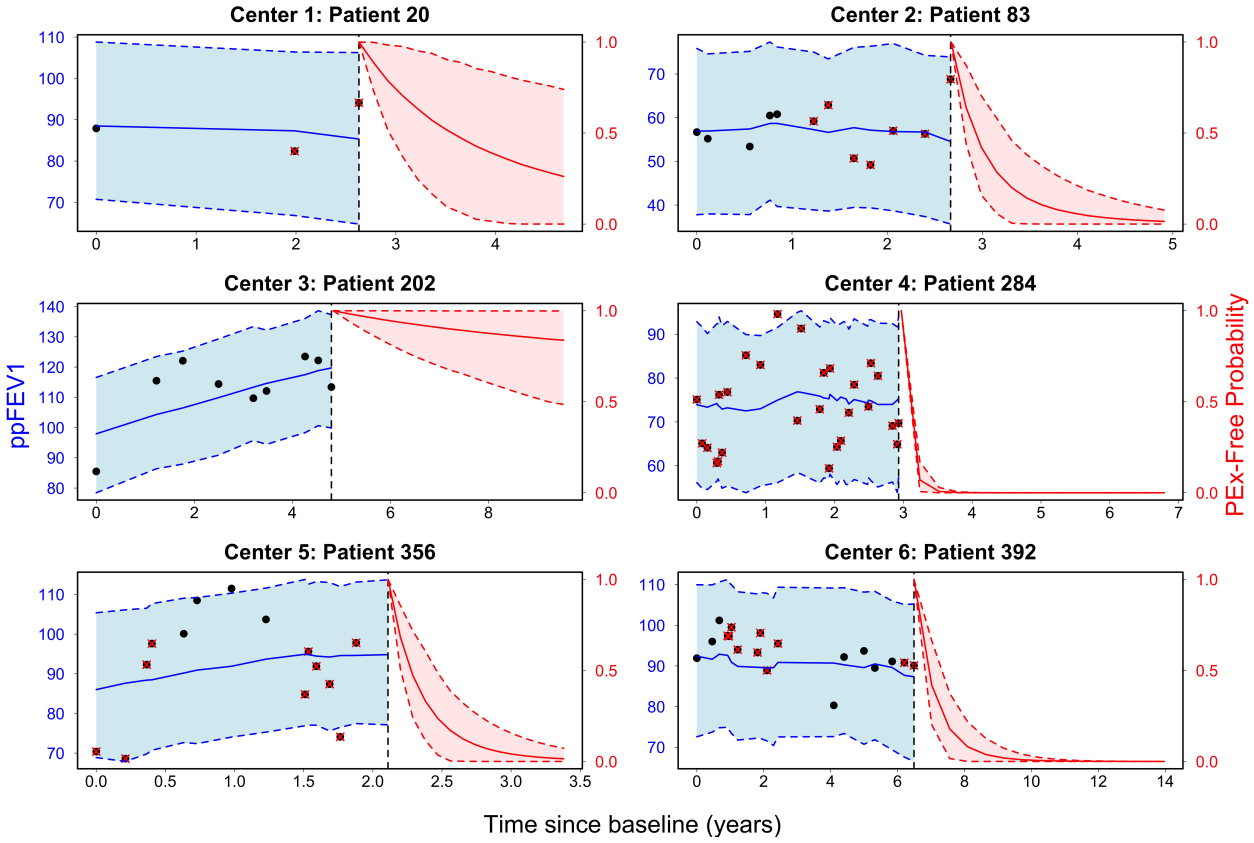

Figure 3.6: Individual predictions against time by centers, including observed ppFEV1 (dot) illustrated by PEx event (red cross) and non-PEx event (black dot), fitted value of ppFEV1 (blue line), prognostic PEx-free probability (red line) and 95% credible interval (band)

The availability of robust software packages is limited. In Chapter 1, we have introduced briefly R packages JMBayes2 and rstanarm. There also exists R package frailtypack (Rondeau et al. (2019)), which allows for joint nested frailty models in the context of the joint modeling for recurrent events with terminal event, for hierarchically clustered data. Notwithstanding that it closely relates to our topic, it adapts maximum penalized likelihood estimation rather than the Bayesian approach. For this sake, we implement our own R code with Stan programs via CmdStanR interface to Stan (Gabry and Češnovar (2022)). To en-

hance rigor and reproducibility of our proposed model, the example codes have been posted in Appendix C.7.

We consider joint model and two-stage approach in this study. Both approaches estimate model parameters using HMC. However, the intrinsic rationale is different. The joint model fits two submodels simultaneously using shared random effects via one HMC, nonetheless, the two-stage method involves two consecutive HMCs, but the model setup (likelihood function) is simpler, thus less computational time (see the comparison in Appendix C.6). As addressed in Barrett et al. (2019), the method of choice depends on the extent of the biases incurred by the two-stage approach and the computational tractability of the joint modeling approach. The two-stage approach is appropriate if the truncation of the observation process is noninformative, however, in our CF study, we believe that the occurrence of PEx will affect the future values of ppFEV1. Therefore, we adapt to a trade-off modeling strategy, which utilizes the two-stage method for model selection and the joint model for Bayesian inference.

This chapter would be considered in the light of some limitations and extensions. To include an additional submodel for a terminal event in the context of the extended joint model might be further investigated. Notwithstanding we present the individual prognostic rationale for in-sample patients, there exists great practical and theoretical interest to investigate out-of-sample predictions in terms of dynamic feature, which was proposed with novelty for univariate joint model in Proust-Lima and Taylor (2017) and Rizopoulos (2011); applied for joint model with recurrent events in Ren et al. (2021) or joint model for re-

current events with a terminal event in Mauguen et al. (2013). Furthermore, with respect to the specific forecast period, it might be interesting to access the predictive Performance at several clinical relevant window widths (e.g., age 18 to 23 or age 25 to 30), which are dependent on frequencies of the next PEx events. Though the predictive probability of next event is welcome, an alternative measure is mean residual life (MRL), which provides the expected time to the next event occurrence (Deep et al. (2020)). The advantage of MRL lies in its clinical interpretation between physicians and patients. We utilize ubiquitous random intercept-slope model for the longitudinal submodel, while it can be replaced by some other novel Gaussian process models, nevertheless, the complexity is unlikely to be warranted. In this chapter, we barely include time-varying covariates for the event submodel, hence, more time-dependent risk factors can be further studied. With respect to various baseline hazard function candidates, alternatives such as bsplines, piecewise-constant are undergoing. Lastly, we investigate the routine diagnostics of joint model, which has not received much attention in the joint modeling literature. Readers who are interested in a thorough discussion about this topic can refer to Rizopoulos (2012b) for some new insights.

Chapter 4

Conclusion and Future Work

To summarize, we have proposed two novel multilevel Bayesian joint models in hierarchically structured CF data, which both present reasonable personalized prediction for the PEx risk. Specifically, in Chapter 2, we treat PEx as the longitudinal binary outcome and interested in its occurrence. Whilst in Chapter 3, we treat PEx as the recurrent outcome so that we can predict the time to the next recurrent event. The different motivation results in different submodel frameworks, nonetheless, we apply center-specific association parameter to both studies, which greatly facilitate the model performance.

The novelty of this dissertation is evident, such as flexible link function with center-specific power parameter, center-specific latent trajectory and baseline hazard function, individual-specific prediction start time and customized applicable Stan programs, particularly designed for the CF study. In addition, this application can be extended for the cases of multivariate joint model or numerous alternative hierarchical data structures.

The future works of this dissertation are of interest to several aspects. For instance, to make our joint model applicable to the public, we are implementing an interactive web app from RShiny (RStudio, Inc (2013)) that can incorporate dynamic individual prediction and some other visualization functions. Besides predictive accuracy, we are motivated to investigate some parallel computation techniques to reduce the cost of intensive computational time. Lastly, we can extend our study to some remedies for left-truncation phenomenon (Król et al. (2016), Piccorelli and Schluchter (2012)) and heavy-tailedness behavior (Asar et al. (2018)).

Bibliography

Andrinopoulou ER, Rizopoulos D, Takkenberg JJ, Lesaffre E (2017) Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. Statistical Methods in Medical Research 26(4):1787–1801, DOI 10.1177/0962280215588340

Andrinopoulou ER, Eilers P, Takkenberg J, Rizopoulos D (2018) Improved dynamic predictions from joint models of longitudinal and survival data with time-varying effects using p-splines. Biom 74:685–693, DOI https://doi-org.uc.idm.oclc.org/10.1111/biom.12814

Andrinopoulou ER, Harhay MO, Ratcliffe SJ, Rizopoulos D (2021) Reflection on modern methods: Dynamic prediction using joint models of longitudinal and time-to-event data. International Journal of Epidemiology 50(5):1731–1743, DOI https://doi.org/10.1093/ije/dyab047

Asar Ö, Ritchie J, Kalra PA, Diggle PJ (2015) Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. International journal of epidemiology 44(1):334–344, DOI https://doi.org/10.1093/ije/dyu262

Asar Ö, Bolin D, Diggle PJ, Wallin J (2018) Linear mixed-effects models for non-gaussian repeated measurement data. DOI 10.48550/ARXIV.1804.02592, URL `https://arxiv.org/abs/1804.02592`

Barrett J, Huille R, Parker R, Yano Y, Griswold M (2019) Estimating the association between blood pressure variability and cardiovascular disease: An application using the aric study. Statistics in Medicine 38:1855–1868, DOI https://doi.org/10.1002/sim.8074

Bayarri MJ, Berger JO (2004) The interplay of bayesian and frequentist analysis. Statistical Science 19(1):58–80, DOI 10.1214/088342304000000116

Betancourt M (2017) Diagnosing biased inference with divergences. `https://mc-stan.org/users/documentation/case-studies/divergences_and_bias.html`, accessed: 2022-5-12

Betancourt M, Girolami M (2015) Current Trends in Bayesian Methodology with Applications, Chapman & Hall/CRC Press, Internet resource, chap Hamiltonian Monte Carlo for Hierarchical Models. First edition

Betancourt MJ, Girolami M (2013) Hamiltonian monte carlo for hierarchical models. DOI 10.48550/ARXIV.1312.0906, URL https://arxiv.org/abs/1312.0906

Brilleman S, Crowther M, Moreno-Betancur M, Buros Novik J, R Wolfe R (2018) Joint longitudinal and time-to-event models via Stan. URL https://github.com/stan-dev/stancon_talks/, stanCon 2018. 10-12 Jan 2018. Pacific Grove, CA, USA.

Brilleman SL, Crowther MJ, Moreno-Betancur M, Novik JB, Dunyak J, Al-Huniti N, Fox R, Hammerbacher J, Wolfe R (2019) Joint longitudinal and time-to-event models for multilevel hierarchical data. Statistical Methods in Medical Research 28(12):3502–3515, DOI 10.1177/0962280218808821

Brombin C, Serio CD, Rancoita PM (2016) Joint modeling of hiv data in multicenter observational studies: A comparison among different approaches. Statistical Methods in Medical Research 25(6):2472–2487, DOI https://doi.org/10.1177/0962280214526192

Brooks S, Gelman A, Jones G, Meng XL (2011) Handbook of Markov Chain Monte Carlo. CRC press

Cox DR, Snell EJ (1968) A general definition of residuals. Journal of the Royal Statistical Society Series B (Methodological) 30(2):248–75, URL http://www.jstor.org/stable/2984505

Deep A, Veeramani D, Zhou D (2020) Event prediction for individual unit based on recurrent event data collected in teleservice systems. IEEE Transactions on Reliability 69(1):216–227, DOI 10.1109/TR.2019.2909471

Diggle P, Sousa I, Asar O (2015) Real-time monitoring of progression towards renal failure in primary care patients. Biostatistics 16(3):522–536, DOI https://doi.org/10.1002/bimj.201900044

Gabry J, Češnovar R (2022) cmdstanr: R Interface to 'CmdStan'. Https://mc-stan.org/cmdstanr/, https://discourse.mc-stan.org

Gabry J, Mahr T (2020) bayesplot: Plotting for bayesian models. URL https://mc-stan.org/bayesplot, r package version 1.7.2

Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85(410):398–409, DOI https://doi.org/10.2307/2289776

Gelman A (2020) Prior choice recommendations. https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations, accessed: Apr 17, 2020

Gelman A, Carlin J (2014) Beyond power calculations: Assessing type s (sign) and type m (magnitude) errors. Perspectives on psychological science : a journal of the Association for Psychological Science 9(6):641–651, DOI https://doi.org/10.1177/1745691614551642

Gelman A, Rubin D (1992) Inference from iterative simulation using multiple sequences. Statistical Science 7(4):457–511, DOI 10.1214/ss/1177011136

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013a) Bayesian Data Analysis, Chapman & Hall/CRC Press, London, chap 4 Hierarchical models, pp 111–148. Third edition

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013b) Bayesian Data Analysis, Chapman & Hall/CRC Press, London, chap 11.4 Inference and Assessing Convergence, p 285. Third edition

Gelman A, Hwang J, Vehtari A (2014) Understanding predictive information criteria for bayesian models. Stat Comput 24:997–1016, DOI https://doi.org/10.1007/s11222-013-9416-2

Geman S, Geman D (1984) Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6(6):721–741, DOI 10.1109/TPAMI.1984.4767596

Geyer CJ (1992) Practical markov chain monte carlo. Statist Sci 7(4):473 – 483, DOI https://doi.org/10.1214/ss/1177011137

Goodrich B, Gabry J, Ali I, Brilleman S (2020) rstanarm: Bayesian applied regression modeling via Stan. URL `https://mc-stan.org/rstanarm`, r package version 2.21.1

Hackenberger BK (2019) Bayes or not bayes, is this the question? Croatian medical journal 60(1):50–52, DOI 10.3325/cmj.2019.60.50

Han J, Slate E, Peña E (2007) Parametric latent class joint model for a longitudinal biomarker and recurrent events. Stat Med 26(29):5285–302, DOI 10.1002/sim.2915

Harhay MO, Au DH, Dell SD, Gould MK, Redline S, Ryerson CJ, Cooke CR (2020) Methodologic guidance and expectations for the development and reporting of prediction models and causal inference studies. Annals of the American Thoracic Society 17(6):679–682, DOI https://doi.org/10.1513/AnnalsATS.202002-141ED

Hastings WK (1970) Monte carlo sampling methods using markov chains and their applications. Biometrika 57(1):97–109, DOI https://doi.org/10.2307/2334940

Heinze G, Dunkler D (2017) Five myths about variable selection. Transplant international : official journal of the European Society for Organ Transplantation 30(1):6–10, DOI https://doi.org/10.1111/tri.12895

Henderson R, Diggle P, Dobson A (2000) Joint modelling of longitudinal measurements and event time data. Biostatistics 1(4):465–480, DOI 10.1093/biostatistics/1.4.465

Hickey G, Philipson P, Jorgensen A, R KD (2018a) Joint models of longitudinal and time-to-event data with more than one event time outcome: A review. Int J Biostat 14(1):3502–3515, DOI 10.1515/ijb-2017-0047

Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R (2016) Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. BMC Med Res Methodol 16(117), DOI https://doi-org.proxy.libraries.uc.edu/10.1186/s12874-016-0212-5

Hickey GL, Philipson P, Jorgensen A (2018b) joinerml: a joint model and software package for time-to-event and multivariate longitudinal outcomes. BMC Medical Research Methodology 18(1), DOI 10.1186/s12874-018-0502-1, URL `https://doi.org/10.1186%2Fs12874-018-0502-1`

Hocking RR (1976) A biometrics invited paper. the analysis and selection of variables in linear regression. Biometrics 32(1):1–49, DOI https://doi.org/10.2307/2529336

Hoffman MD, Gelman A (2011) The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. DOI 10.48550/ARXIV.1111.4246, URL `https://arxiv.org/abs/1111.4246`

Horrocks J, van Den Heuvel MJ (2009) Prediction of pregnancy: a joint model for longitudinal and binary data. Bayesian Anal 4(3):523–538, DOI https://doi.org/10.1214/09-BA419

Ibrahim J, Chu H, Chen L (2010) Basic concepts and methods for joint models of longitudinal and survival data. J Clin Oncol 28(16):2796–2801, DOI https://doi.org/10.1200/JCO.2009.25.0654

Jiang X, Dey DK, Prunier R, Wilson AM, Holsinger KE (2013) A new class of flexible link functions with application to species co-occurrence in cape floristic region. Ann Appl Stat 7(4):2180–2204, DOI https://doi.org/10.1214/13-AOAS663

Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53(282):457–481

Kim S, Zeng D, Chambless L, Li Y (2012) Joint models of longitudinal data and recurrent events with informative terminal event. Stat Biosci 4(2):262–281, DOI 10.1007/s12561-012-9061-x

Knapp E, Fink A, Goss C, Sewall A, Ostrenga J, Dowd C, Elbert A, Petren K, BC M (2016) The cystic fibrosis foundation patient registry. design and methods of a national observational disease registry. Ann Am Thorac Soc 13(7):1173–9, DOI 10.1513/AnnalsATS.201511-781OC

Król A, Ferrer L, Pignon JP, Proust-Lima C, Ducreux M, Bouché O, Michiels S, Rondeau V (2016) Joint model for left-censored longitudinal data, recurrent events and terminal event: Predictive abilities of tumor burden for cancer evolution with application to the ffcd 2000-05 trial. Biometrics 72(3):907–916, DOI 10.1111/biom.12490

Laird N, Ware J (1982) Random-effects models for longitudinal data. Biometrics 38(4):963–974, pMID: 7168798

Laurie D (1997) Calculation of gauss-kronrod quadrature rules. Math Comput 66(219):1133–45

Lewandowski D, Dorota K, Harry J (2009) Generating random correlation matrices based on vines and extended onion method. Journal of Multivariate Analysis 100:1989–2001

Li D, Keogh R, Clancy J, Szczesniak R (2017) Flexible semiparametric joint modeling:an application to estimate individual lung function decline and risk of pulmonary exacerbations in cystic fibrosis. Emerg Themes Epidemiol 14(13), DOI 10.1186/s12982-017-0067-1

Liu L, Huang X, O'Quigley J (2008) Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. Biometrics 64(3):950–958, DOI 10.1111/j.1541-0420.2007.00954.x

Luo S (2015) A bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time. Statistics in medicine 33(4):580–594, DOI https://doi.org/10.1002/sim.5956

Luo S, Wang J (2014) Bayesian hierarchical model for multiple repeated measures and survival data: an application to parkinson's disease. Stat Med 33(24):4279–4291, DOI 10.1002/sim.6228

Mauguen A, Rachet B, Mathoulin-Pélissier S, MacGrogan G, Laurent A, Rondeau V (2013) Dynamic prediction of risk of death using history of cancer recurrences in joint frailty models. Statist Med 32:5366–5380, DOI https://doi-org.uc.idm.oclc.org/10.1002/sim.5980

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. Journal of Chemical Physics 21:1087–1092, DOI https://doi.org/10.1063/1.1699114

Musoro J, Geskus R, Zwinderman A (2015) A joint model for repeated events of different types and multiple longitudinal outcomes with application to a follow-up study of patients after kidney transplant. Biom J 57(2):185–200, DOI 10.1002/bimj.201300167

Neal R (2011) MCMC Using Hamiltonian Dynamics, Chapman Hall/CRC, pp 116–62

Philipson P, Sousa I, Diggle PJ, Williamson P, Kolamunnage-Dona R, Henderson R, Hickey GL (2018) joineR: Joint Modelling of Repeated Measurements and Time-to-Event Data. URL `https://github.com/graemeleehickey/joineR/`, r package version 1.2.6

Piccorelli A, Schluchter M (2012) Jointly modeling the relationship between longitudinal and survival data subject to left truncation with applications to cystic fibrosis. Stat Med 31(29):3931–3945, DOI 10.1002/sim.5469

Proust-Lima C, Taylor JM (2017) Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. Biostatistics 10(3):535–549, DOI https://doi.org/10.1093/biostatistics/kxp009

Proust-Lima C, Séne M, Taylor JM, Jacqmin-Gadda H (2014) Joint latent class models for longitudinal and time-to-event data: a review. Statistical methods in medical research 23(1):74–90, DOI https://doi.org/10.1177/0962280212445839

Proust-Lima C, Philipps V, Diakite A, Liquet B (2022) lcmm: Extended Mixed Models Using Latent Classes and Latent Processes. URL `https://cran.r-project.org/package=lcmm`, r package version: 1.9.5

Quanjer PH, Stanojevic S, Cole TJ, Baur X, L HG, Culver BH, Enright PL, Hankinson JL, Ip MS, Zheng J, Stocks J, the ERS Global Lung Function Initiative (2012) Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. European Respiratory Society 40(6):1324–1343, DOI https://doi.org/10.1183/09031936.00080312

R Core Team (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL `https://www.R-project.org/`

Ren X, Wang J, Luo S (2021) Dynamic prediction using joint models of longitudinal and recurrent event data: a bayesian perspective. Biostatistics & Epidemiology 5(2):250–266, DOI 10.1080/24709360.2019.1693198

Rizopoulos D (2010) Jm: An r package for the joint modelling of longitudinal and time-to-event data. Journal of Statistical Software 35(9):1–33, DOI https://doi.org/10.18637/jss.v035.i09

Rizopoulos D (2011) Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. Biometrics 67(3):819–829, DOI 10.1111/j.1541-0420.2010.01546.x

Rizopoulos D (2012a) Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. Chapman and Hall/CRC, Boca Raton, FL

Rizopoulos D (2012b) Joint Models for Longitudinal and Time-to-Event Data: With Applications in R, Chapman and Hall/CRC, Boca Raton, FL, chap 6, pp 145–169

Rizopoulos D (2012c) Stratified Relative Risk Models, Chapman and Hall/CRC, Boca Raton, FL, chap 5.3, pp 119–122

Rizopoulos D (2016) The r package jmbayes for fitting joint models for longitudinal and time-to-event data using mcmc. journal of statistical software. Journal of Statistical Software 72(7):1–46, DOI https://doi.org/10.18637/jss.v072.i07

Rizopoulos D, Hatfield LA, Carlin BP, Takkenberg JJM (2014) Combining dynamic predictions from joint models for longitudinal and time-to-event data using bayesian model averaging. Journal of the American Statistical Association 109(508):1385–1397, DOI 10.1080/01621459.2014.931236

Rizopoulos D, Papageorgiou G, Miranda Afonso P (2022) JMbayes2: Extended Joint Models for Longitudinal and Time-to-Event Data. Https://drizopoulos.github.io/JMbayes2/, https://github.com/drizopoulos/JMbayes2

Rondeau V, Mazroui Y, Gonzalez JR (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(4):1–28, URL `https://www.jstatsoft.org/v47/i04/`

Rondeau V, Gonzalez JR, Mazroui Y, Mauguen A, Diakite A, Laurent A, Lopez M, Król A, Sofeu CL (2019) frailtypack: General Frailty Models: Shared, Joint and Nested Frailty Models with Prediction; Evaluation of Failure-Time Surrogate Endpoints. URL `https://CRAN.R-project.org/package=frailtypack`, R package version 3.0.3

RStudio, Inc (2013) Easy web applications in R. URL: `http://www.rstudio.com/shiny/`

Rubin DB (1981) Estimation in parallel randomized experiments. Journal of Educational and Behavioral Statistics 6(4):377–401, URL `https://EconPapers.repec.org/RePEc:sae:jedbes:v:6:y:1981:i:4:p:377-401`

Shen Y, Huang H, Guan Y (2016) A conditional estimating equation approach for recurrent event data with additional longitudinal information. Stat Med 35(24):4306–4319, DOI 10.1002/sim.7001

Smedinga H, Steyerberg EW, Beukers W, Klaveren Dv, Zwarthoff EC, Vergouwe Y (2017) Prediction of multiple recurrent events: A comparison of extended cox models in bladder cancer. American Journal of Epidemiology 186(5):612–623, DOI 10.1093/aje/kwx133

Sorensen T, Hohenstein S, Vasishth S (2016) Bayesian linear mixed models using stan: A tutorial for psychologists, linguists, and cognitive scientists. the quantitative methods for psychology. The Quantitative Methods for Psychology 12(3):175–200, DOI 10.20982/tqmp.12.3.p175

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit. J Roy Stat Soc B 64(4):583–639, DOI https://doi.org/10.1111/1467-9868.00353

Stan Development Team (2011-2019) Stan Modeling Language Users Guide and Reference Manual. URL `https://mc-stan.org`, version 2.29

Stan Development Team (2020) Rstan: the R interface to Stan. R Foundation for Statistical Computing, Vienna, Austria, URL `http://mc-stan.org/`, r package version 2.21.2

Su W, Wang X, Szczesniak RD (2020) Flexible link functions in a joint hierarchical gaussian process model. Biometrics pp 1–11, DOI https://doi.org/10.1111/biom.13291

Su W, Wang X, Szczesniak RD (2021) Risk factor identification in cystic fibrosis by flexible hierarchical joint models. Statistical Methods in Medical Research 30(1):244–260, DOI https://doi.org/10.1177/0962280220950369

Szczesniak R, Li D, Su W, Brokamp C, Pestian J, Seid M, Clancy J (2017) Phenotypes of rapid cystic fibrosis lung disease progression during adolescence and young adulthood. American Journal of Respiratory & Critical Care Medicine 196:471–478

Szczesniak R, Su W, Brokamp C, Ruth H, Pestian J, Seid M, Diggle P, Clancy J (2020) Dynamic predictive probabilities to monitor rapid cystic fibrosis disease progression. Statistics in Medicine 39:740–756, DOI https://doi.org/10.1002/sim.8443

Szczesniak RD, McPhail GL, Duan LL, Macaluso M, Amin RS, Clancy JP (2013) A semiparametric approach to estimate rapid lung function decline in cystic fibrosis. Annals of Epidemiology 23(12):771–777, DOI https://doi.org/10.1016/j.annepidem.2013.08.009.

Tanner MA, Wong WH (1987) The calculation of posterior distributions by data augmentation (with discussion). Journal of the American Statistical Association 82:528–550, DOI https://www.jstor.org/stable/2289457

Taylor-Robinson D, Whitehead M, Diderichsen F, Olesen H, Pressler T, Smyth R, Diggle P (2012) Understanding the natural progression in %fev decline in patients with cystic fibrosis: a longitudinal study. Thorax 67:860–866

Tierney L (1994) Markov chains for exploring posterior distributions. Ann Statist 22(4):1701 – 1728, DOI https://doi.org/10.1214/aos/1176325750

Tsiatis AA, Davidian M (2004) Joint modeling of longitudinal and time-to-event data: An overview. Statistica Sinica 14(3):809–834, URL https://www.jstor.org/stable/24307417

Vehtari A, Gelman A, Gabry J (2017) Practical bayesian model evaluation using leave-one-out cross-validation and waic. Stat Comput 27:1413–1432, DOI 10.1007/s11222-016-9696-4

Vehtari A, Gabry J, Magnusson M, Yao Y, Bürkner PC, Paananen T, Gelman A (2020) loo: Efficient leave-one-out cross-validation and waic for bayesian models. URL https://mc-stan.org/loo/, r package version 2.4.1

Voeten CC (2021) buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression. URL https://CRAN.R-project.org/package=buildmer, r package version 1.8

Wang X, Dey D (2010) Generalized extreme value regression for binary response data: an application to b2b electronic payments system adoption. The Annals of Applied Statistics, 4:2000–2023

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

Wulfsohn MS, Tsiatis AA (1997) A joint model for survival and longitudinal data measured with error. Biometrics 53(1):330–339, URL https://www.jstor.org/stable/2533118

Appendix A

Appendix for Chapter 1

A.1 Example Code

A.1.1 Example data

```
# eight\_schools.data.R

J <- 8

y <- c(28, 8, -3, 7, -1, 1, 18, 12)

sigma <- c(15, 10, 16, 11, 9, 11, 10, 18)
```

A.1.2 Stan program

```
# eight_schools_cp.stan
```

```
data {
  int<lower=0> J;
  real y[J];
  real<lower=0> sigma[J];
}

parameters {
  real mu;
  real<lower=0> tau;
  real theta[J];
}

model {
  mu ~ normal(0, 5);
  tau ~ cauchy(0, 5);
  theta ~ normal(mu, tau);
  y ~ normal(theta, sigma);
}

# eight_schools_ncp.stan
```

```
data {
  int<lower=0> J;
  real y[J];
  real<lower=0> sigma[J];
}

parameters {
  real mu;
  real<lower=0> tau;
  real theta_tilde[J];
}

transformed parameters{
  real theta[J];
  for (j in 1:J)
    theta[j]=mu+tau*theta_tilde[j];
}

model {
  mu ~ normal(0, 5);
  tau ~ cauchy(0, 5);
```

```
  theta_tilde ~ normal(0, 1);
  y ~ normal(theta, sigma);
}
```

A.1.3 R code

```
# Setup


library(gridExtra) #v2.3
library(bayesplot) #v1.8.1
library(dplyr) #v1.0.5
library(cmdstanr) #v0.4.0
library(rstan) #v2.21.1
options(mc.cores = parallel::detectCores())
rstan_options(auto_write = TRUE)
input_data <- read_rdump('eight_schools.data.R')


# Fit


##Default: adapt_delta=0.8, max_treedepth=10, iter=2000, warmup=1000
###rstan
```

```
    fit_cp <- rstan::stan('eight_schools_cp.stan', data = input_data,
                          iter = 2000, warmup = 1000, chains = 2, seed = 2022)

    rstan::get_elapsed_time(fit_cp)

    fit_ncp <- rstan::stan(file = 'eight_schools_ncp.stan', data = input_data,
                          iter = 2000, warmup = 1000, chains = 2, seed = 2022)

###cmdstanr
  mod.cp <- cmdstanr::cmdstan_model('eight_schools_cp.stan')
  fit_cp_cmd <- mod.cp$sample(data = input_data, chains = 2,
                        save_warmup = FALSE, parallel_chains = 2, refresh = 500, seed=2022)

  mod.ncp <- cmdstanr::cmdstan_model('eight_schools_ncp.stan')
  fit_ncp_cmd <- mod.ncp$sample(data = input_data, chains = 2,
                        save_warmup = FALSE, parallel_chains = 2, refresh = 500,
                        seed=2022)

# Results
##rstan
tau_cp0 <- as.data.frame(extract(fit_cp, par='tau', permuted=TRUE))
```

```
tau_cp <- tau_cp0 %>%
          rename(tau.cp=tau) %>%
          mutate(log.tau.cp=log(tau.cp))

tau_ncp0 <- as.data.frame(extract(fit_ncp, par='tau', permuted=TRUE))
tau_ncp <- tau_ncp0 %>%
           rename(tau.ncp=tau) %>%
           mutate(log.tau.ncp=log(tau.ncp))

tau_all <- cbind(tau_cp,tau_ncp) %>%
            mutate(iter=1:2000,
                   chain=rep(1:2,each=1000))

tau_all$mean.cp <- sapply(tau_all$iter, function(n) mean(tau_all$log.tau.cp[1:n]))
tau_all$mean.ncp <- sapply(tau_all$iter, function(n) mean(tau_all$log.tau.ncp[1:n]))

tau_all$div.cp <- c(get_sampler_params(fit_cp, inc_warmup=FALSE)[[1]][,'divergent__'],
                    get_sampler_params(fit_cp, inc_warmup=FALSE)[[2]][,'divergent__'])

tau_all$div.ncp <- c(get_sampler_params(fit_ncp, inc_warmup=FALSE)[[1]][,'divergent__'],
                     get_sampler_params(fit_ncp, inc_warmup=FALSE)[[2]][,'divergent__'])
```

```
##cmdstanr
tau_cp_cmd <- fit_cp_cmd$draws('tau',format='df') %>%
              rename(chain=.chain,
                     tau.cp=tau) %>%
              mutate(log.tau.cp=log(tau.cp),
                     iter=1:2000)

tau_ncp_cmd <- fit_ncp_cmd$draws('tau',format='df') %>%
                rename(chain=.chain,
                       tau.ncp=tau) %>%
                mutate(log.tau.ncp=log(tau.ncp),
                       iter=1:2000)

tau_all_cmd <- tau_cp_cmd %>%
              left_join(tau_ncp_cmd, by=c('chain','iter','.draw'))


mean.cp <- sapply(tau_all_cmd$iter, function(n) mean(tau_all_cmd$log.tau.cp[1:n]))
tau_all_cmd$mean.cp <- mean.cp

mean.ncp <- sapply(tau_all_cmd$iter,function(n) mean(tau_all_cmd$log.tau.ncp[1:n]))
```

```
tau_all_cmd$mean.ncp <-mean.ncp


diag.cp <- fit_cp_cmd$sampler_diagnostics(format = "df")

diag.ncp <- fit_ncp_cmd$sampler_diagnostics(format = "df")


tau_all_cmd$div.cp <- diag.cp$divergent__

tau_all_cmd$div.ncp <- diag.ncp$divergent__


sum(tau_all$div.cp)/2000 #0.0115

sum(tau_all$div.ncp)/2000 #0

sum(tau_all_cmd$div.cp)/2000 #0.04

sum(tau_all_cmd$div.ncp)/2000 # 0


#Plot

light.orange='#E69F00'

orange='#D55E00'

light.blue='#56B4E9'

blue='#0072B2'


#Figure 1.1

{
```

```
jpeg("Chp1_mean_tau.jpg", width = 350, height = 300, units='mm', res = 300)
par(cex.lab=2, cex.axis=1.5)
par(mar = c(4, 5, 0.5, 0.5))
plot(NA, xlim=c(1,2000), ylim=c(0, 2),xlab="Iteration", ylab="mean of log(tau)")
points(tau_all$iter,tau_all$mean.cp,col=light.orange, pch=16,cex=2)
points(tau_all$iter,tau_all$mean.ncp,col=orange, pch=16,cex=2)
points(tau_all_cmd$iter,tau_all_cmd$mean.cp,col=light.blue, pch=16,cex=2)
points(tau_all_cmd$iter,tau_all_cmd$mean.ncp,col=blue, pch=16,cex=2)
abline(h=0.7657852, col='#009E73', lty="dashed", lwd=5)

div_iter_cp <- tau_all$iter[which(tau_all$div.cp==1)]
div_mean_cp <- tau_all$mean.cp[which(tau_all$div.cp==1)]
div_iter_cp_cmd <- tau_all_cmd$iter[which(tau_all_cmd$div.cp==1)]
div_mean_cp_cmd <- tau_all_cmd$mean.cp[which(tau_all_cmd$div.cp==1)]

points(div_iter, div_mean ,col="red", pch=4, lwd=2,cex=2)
points(div_iter_cp_cmd, div_mean_cp_cmd ,col="red", pch=4, lwd=2, cex=2)

box(lwd=3)
legend("bottomright",
       c("rstan: centered", "rstan: non-centered",
```

```
           "cmdstanr: centered", 'cmdstanr: non-centered','divergence','true'),
         pch=c(rep(16,4), 4, NA),
         lty=c(NA,NA,NA,NA,NA,'dashed'),
         lwd=c(NA,NA,NA,NA,2,3),
         col=c(light.orange, orange, light.blue, blue, 'red', "#009E73"),
         cex=1.5, bty="n")

  dev.off()
}

#Figure 1.2
color_scheme_set("brightblue")
tt1 <- mean(rstan::get_elapsed_time(fit_cp))
tt2 <- mean(rstan::get_elapsed_time(fit_ncp))
tt3 <- mean(fit_cp_cmd$time()$chains$total)
tt4 <- mean(fit_ncp_cmd$time()$chains$total)

{
pp1 <- bayesplot::mcmc_trace(tau_all,  pars = c("log.tau.cp")) +
  labs(title = paste0("rstan: centered (",round(tt1,2),'s)')) +
    theme(plot.title = element_text(hjust = 0.5,face='bold',size=20),
```

```
          legend.title = element_text(size=12),
          legend.text = element_text(size=12),
          axis.text.x = element_text(size=15),
          axis.text.y = element_text(size=15),
          axis.title.x = element_text(size=15),
          axis.title.y = element_text(size=15))

pp2 <- bayesplot::mcmc_trace(tau_all,  pars = c("log.tau.ncp")) +
  labs(title = paste0("rstan: non-centered (",round(tt2,2),'s)')) +
  theme(plot.title = element_text(hjust = 0.5,face='bold',size=20),
        legend.title = element_text(size=12),
        legend.text = element_text(size=12),
        axis.text.x = element_text(size=15),
        axis.text.y = element_text(size=15),
        axis.title.x = element_text(size=15),
        axis.title.y = element_text(size=15))

pp3 <- bayesplot::mcmc_trace(tau_all_cmd,  pars = c("log.tau.cp"))+
  labs(title = paste0("cmdstanr: centered (",round(tt3,2),'s)')) +
  theme(plot.title = element_text(hjust = 0.5,face='bold',size=20),
        legend.title = element_text(size=12),
```

```
        legend.text = element_text(size=12),
        axis.text.x = element_text(size=15),
        axis.text.y = element_text(size=15),
        axis.title.x = element_text(size=15),
        axis.title.y = element_text(size=15))

pp4 <- bayesplot::mcmc_trace(tau_all_cmd,  pars = c("log.tau.ncp"))+
  labs(title = paste0("cmdstanr: non-centered (",round(tt4,2),'s)')) +
  theme(plot.title = element_text(hjust = 0.5,face='bold',size=20),
        legend.title = element_text(size=12),
        legend.text = element_text(size=12),
        axis.text.x = element_text(size=15),
        axis.text.y = element_text(size=15),
        axis.title.x = element_text(size=15),
        axis.title.y = element_text(size=15))

  pp.all=gridExtra::grid.arrange(pp1, pp2, pp3, pp4, nrow = 2, ncol=2)
  ggsave('Chp1_trace_tau.jpg', plot = pp.all, scale = 1, width = 35,
         height = 30, units = c("cm"), dpi = 300)
}
```

A.2 System and versions

Table A.1: Processing system

	Simulated data
Platform	x86_64-apple-darwin17.0 (64-bit)
Running under	macOS Big Sur 10.16
R version	4.0.5 (2021-03-31)
bayesplot	v1.8.1
dplyr	v1.0.5
gridExtra	v2.3
cmdstanr	v0.4.0
CmdStan	v2.29.2
rstan	v2.21.1

Appendix B

Appendix for Chapter 2

B.1 Symmetric Power Link Family

Jiang et al. (2013) demonstrated that

$$F_{sp}(x;r)\begin{cases} \text{local positive (right) skewness if } 0 < r < 1 \\ \text{symmetric if } r = 1 \\ \text{local negative(left) skewness if } r > 1 \end{cases}$$

This statement can be verified given symmetric x in Figure 2.1 from the manuscript however, the opposite side cannot be true. In other words, we may observe right or left skewness for responses (e.g., more or less 1's), however, this cannot guarantee the expected range of power parameter. We present a simulation study based on 1000 replicates in Table B.1 and note that response sknewss (1's %) varies across power parameter r if x is asymmetric.

Table B.1: Relationship between response skewness (1's %) and r given different range of x

	SPLOGIT			SPEP		
x	$r = 0.5$	$r = 1$	$r = 2$	$r = 0.5$	$r = 1$	$r = 2$
$[-0.625, 0]$	59	42	20	53	37	16
$[0, 0.625]$	80	58	41	84	63	47
$[-0.625, 0.625]$	70	50	30	68	50	32
$[-5, 0]$	18	14	4	14	10	3
$[0, 5]$	96	86	82	97	90	86
$[-5, 5]$	57	50	43	56	50	44

B.2 Data Clean

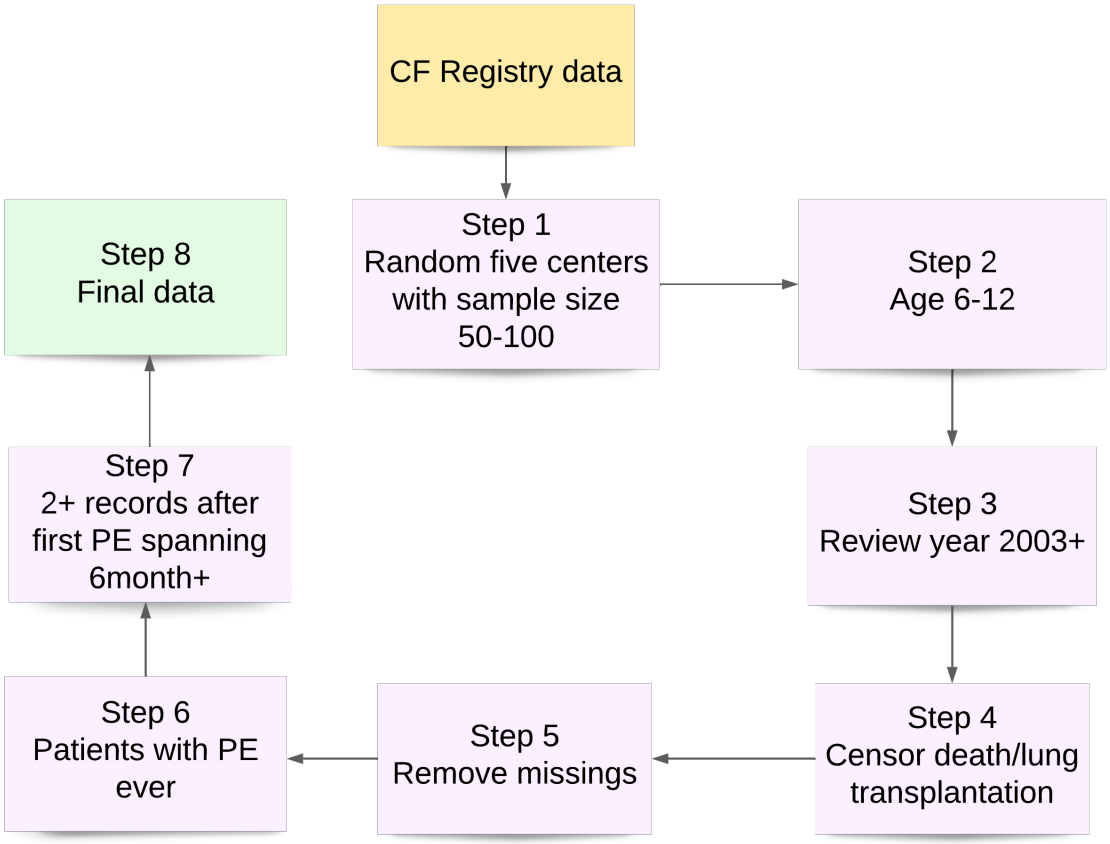

Figure B.1: Data cleaning process

Table B.2: Clinical and demographic summary for CF data

	Training Cohort (N=302)	Testing Cohort (N=79)	Total (N=381)
Center ID			
1	53 (17.5%)	14 (17.7%)	67 (17.6%)
2	81 (26.8%)	21 (26.6%)	102 (26.8%)
3	56 (18.5%)	15 (19.0%)	71 (18.6%)
4	77 (25.5%)	20 (25.3%)	97 (25.5%)
5	35 (11.6%)	9 (11.4%)	44 (11.5%)
Genotpye (F508del)			
Homozygous	176 (58.3%)	48 (60.8%)	224 (58.8%)
Heterozygous	99 (32.8%)	26 (32.9%)	125 (32.8%)
Neither/unknown	27 (8.9%)	5 (6.3%)	32 (8.4%)
Age at baseline (years)			
Mean; Median (Min - Max)	7.94; 7.65 (6.0 - 11.4)	7.76; 7.34 (6.0 - 11.1)	7.90; 7.56 (6.0 - 11.4)
ppFEV1 at baseline			
Mean; Median (Min - Max)	86.7; 88.2 (32.1 - 148)	79.1; 79.3 (30.0 - 126)	85.1; 87.3 (30.0 - 148)
BMI percentile			
Mean; Median (Min - Max)	49.8; 51.0 (0.08 - 98.9)	48.3; 49.2 (0.07 - 98.5)	49.5; 50.4 (0.07 - 98.9)
Insurance			
At baseline	207 (68.5%)	53 (67.1%)	260 (68.2%)
Ever during follow up	246 (81.5%)	65 (82.3%)	311 (81.6%)
Pseudomonas aeruginosa (pa)			
At baseline	52 (17.2%)	11 (13.9%)	63 (16.5%)
Ever during follow up	206 (68.2%)	50 (63.3%)	256 (67.2%)
Methicillin-resistant Staphylococcus aureus (MRSA)			
At baseline	52 (17.2%)	11 (13.9%)	63 (16.5%)
Ever during follow up	206 (68.2%)	50 (63.3%)	256 (67.2%)
Impaired CF-related diabetes mellitus (CFRD)			
At baseline	0 (0%) 0 (0%) 0 (0%)		
Ever during follow up	33 (10.9%)	10 (12.7%)	43 (11.3%)
Enzymes use			
At baseline	175 (57.9%)	44 (55.7%)	219 (57.5%)
Ever during follow up	294 (97.4%)	75 (94.9%)	369 (96.9%)
Pulmonary Exacerbation (PE)			
At baseline	302 (100%)	79 (100%)	381 (100%)
Ever during follow up	296 (98.0%)	77 (97.5%)	373 (97.9%)

B.3 Convergence Diagnostics

Figure B.2 shows that both chains explore the similar region of parameter values. We expect autocorrelation function (ACF) to drop quickly to zero with increasing lag because positive autocorrelation means the chain tends to stay in the same area between iterations. All parameters meet this expectation after lags > 10 in Figure B.3. The larger the ratio of N_{eff} to N the better (Gelman et al. (2013b)), a useful heuristic is to worry about any ratio less than 0.1 (Stan Development Team (2011-2019)), which is not observed in Figure B.4.

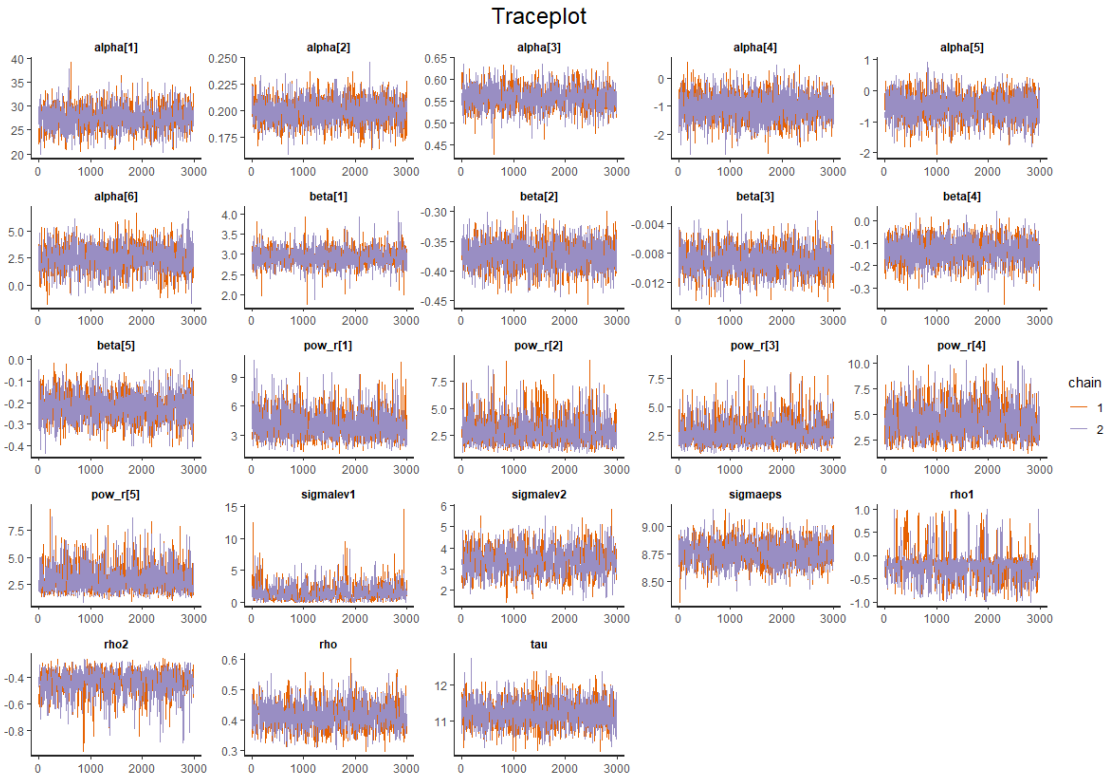

Figure B.2: Traceplot for spep-JM$_4$

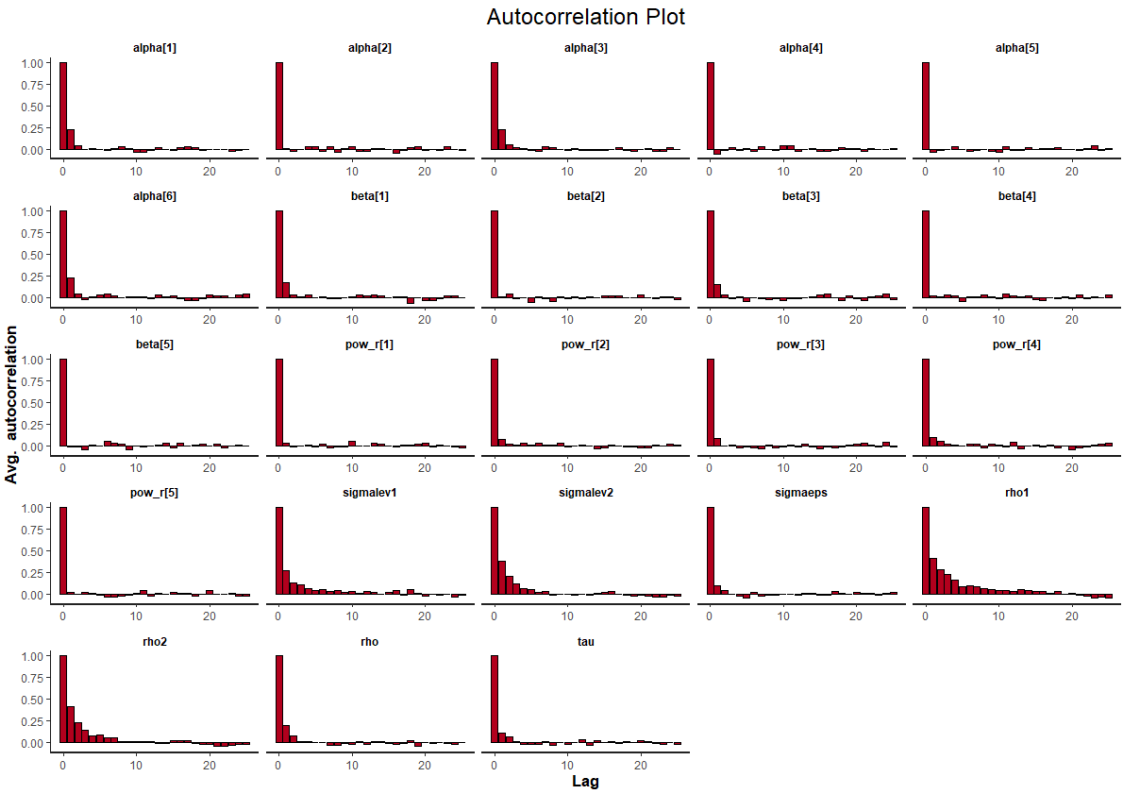

Figure B.3: Autocorrelation plot for spep-JM_4

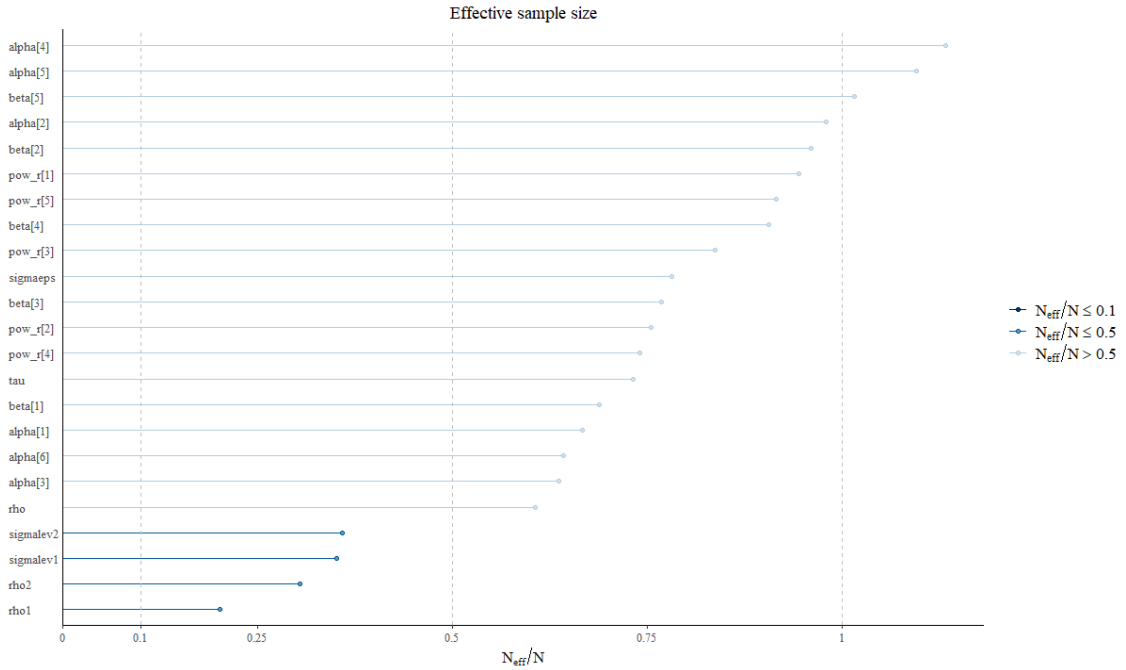

Figure B.4: Effective sample size plot for spep-JM$_4$

B.4 Residual Diagnostics

To access residual diagnostics from the longitudinal submodel, we adapt to the method addressed in Diggle et al. (2015), Szczesniak et al. (2020). The standardized empirical marginal residual is given by

$$\boldsymbol{r}_{li}^{*} = \boldsymbol{S}_{li}^{-1} \cdot \boldsymbol{r}_{li} \tag{B.1}$$

where $\boldsymbol{r}_{li} = \boldsymbol{Y}_{li} - \boldsymbol{X}_{li}\boldsymbol{\alpha}$ denotes a vector of raw residuals. Let $\widehat{\boldsymbol{V}}_{li} = (\hat{\sigma}_b^2 + \hat{\sigma}_u^2)\boldsymbol{J}_{li} + \hat{\tau}^2\boldsymbol{R}_{li} + \hat{\sigma}^2\boldsymbol{I}_{li}$ be the estimated variance-covariance matrix for $\boldsymbol{Y}_{li}$ and $\boldsymbol{J}_{li}$, $\boldsymbol{I}_{li}$ and $\boldsymbol{R}_{li}$ are defined in the

Section 2.2.5. Decompose $\boldsymbol{V}_{li}$ by lower triangular matrix $\boldsymbol{S}_{li}$ such that $\boldsymbol{V}_{li} = \boldsymbol{S}_{li}\boldsymbol{S}_{li}^T$ to achieve $\boldsymbol{S}_{li}^{-1}$.

No clear pattern or trend is detected from Figure B.5. The zero mean of standardized residuals (upper-left & right panel), symmetric bell-shaped distribution (lower-left panel) corroborate model assumptions. The quantile-quantile plot implies a slightly heavier tails than the standard normal distribution.

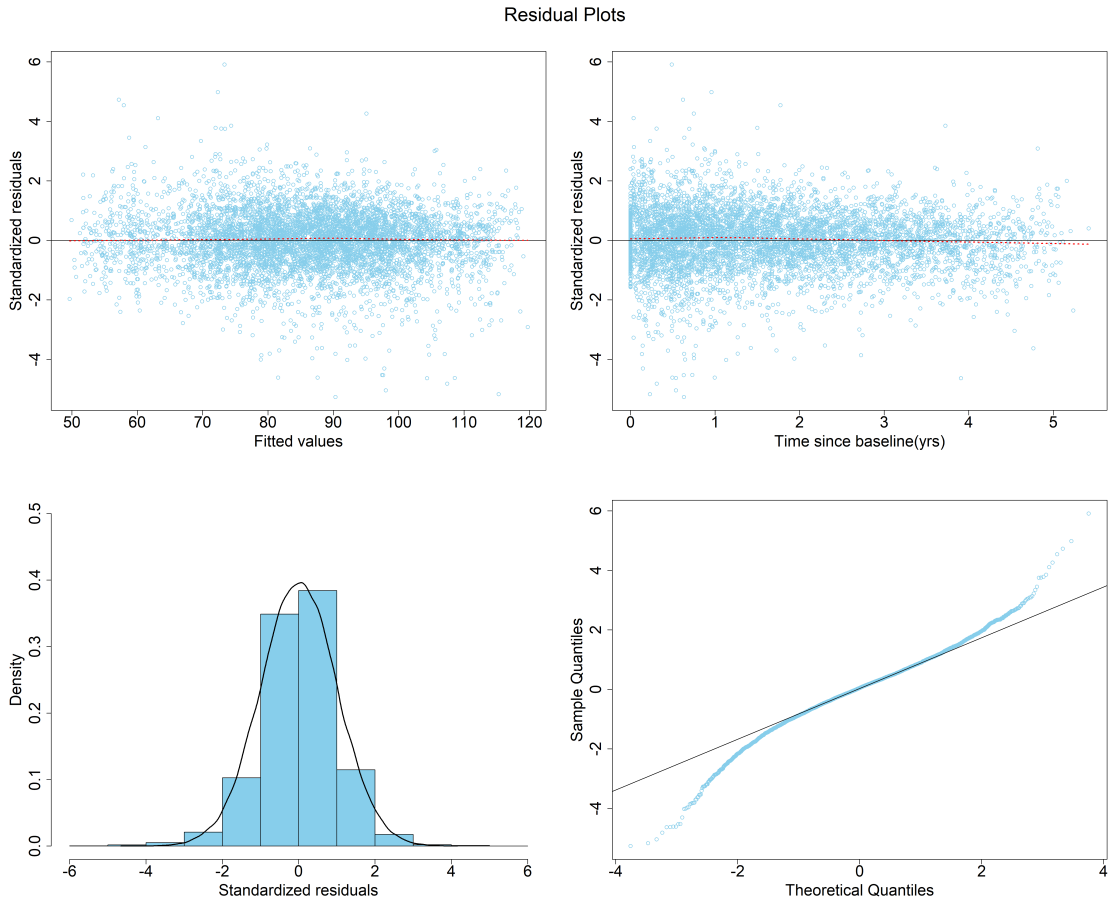

Figure B.5: Diagnostics for standardized residuals from spep-JM4 based on Training Cohort: residuals vs. fitted values (upper left), residuals vs. time variable (upper right), histogram with standard normal density overlay (lower left) and quantile-quantile plot (lower right); LOWESS fitted curves (red dashed lines)

B.5 Time and System

Table B.3: Elapsed time for Simulation A with 50 replicates via CmdStanr

Model	Mac (mins/rep)	PC (mins/rep)
logit-JM	0.24	0.54
probit-JM	0.29	0.68
cloglog-JM	0.29	0.74
splogit-JM	0.39	1.14

Model in boldface: true model; mins=minutes; rep=replicate

Table B.4: Elapsed time for Simulation B with 50 replicates via RStan

Model	Link	Total time (hrs)	PC (hrs/rep)
Misspecified (JM_1)	splogit	9.88	0.20
	spep	7.81	0.16
No center-index (JM_2)	splogit	12.71	0.25
	spep	10.93	0.22
No covariance (JM_3)	splogit	9.51	0.19
	spep	8.88	0.18
Proposed (JM_4)	**splogit**	30.47	0.61
	spep	34.75	0.70

Model in boldface: true model; hrs=hours; rep=replicate

Table B.5: Elapsed time for motivating data via RStan

spep-Model	Total time (hrs)	Post-warmup iters
Misspecified (JM_1)	1.43	5000
No center-index (JM_2)	2.86	8000
No covariance (JM_3)	7.62	8000
Proposed (JM_4)	13.70	5000

Model in boldface: optimal model; hrs=hours; iters=iterations

Table B.6: Processing system and versions

	Mac	**PC**	**BMI**
Platform	x86_64-apple-darwin17.0 (64-bit)	x86_64-w64-mingw32/x64 (64-bit)	x86_64-pc-linux-gnu
Running under	macOS Big Sur 10.16	Windows 10 x64 (build 19043)	x86_64, linux-gnu
R version	4.0.5 (2021-03-31)	4.0.2 (2020-06-22)	3.6.1 (2019-07-05)
CmdStan	v2.28.2	v2.29.1	-
cmdstanr	v0.4.0	v0.5.0	-
rstan	-	-	2.19.2

Note: i). Simulation A: Mac & PC; ii) Simulation B & Real data: Biomedical Informatics (BMI) at CCHMC

B.6 Example Code

We attach an example code for illustration purpose. For the complete code package, please refer to my Github.

B.6.1 Stan Program

```
/*************************************************/
// Purpose: Fit proposed spep-JM4 in the manuscript
// Author: Copyright (C) 2021, 2022 Grace C. Zhou
// Date: Apr., 2021
/*************************************************/

functions {

  //Define exponential power CDF expression

  real  spep_cdf(real x, real pow_r){

    if(pow_r <= 1){
```

```
                return(pow(double_exponential_cdf(x/pow_r, 0, 1), pow_r));

        } else {

                return(1-pow(double_exponential_cdf(-pow_r * x, 0, 1),
                ↪  (1/pow_r)));

         }
  }
}

data {

    int<lower=1> Nobs; //num obs
    int<lower=1> N; //num patients

    int<lower=1> start_pos[N+1]; //starting point index
    int<lower=1> T[N]; //number obs per patient
    vector[Nobs] t; //time since first PE

    int<lower=1> NpredsX; //num LME predictors
    int<lower=1> NpredsV; //num GLMM predictors
    row_vector[NpredsX] X[Nobs]; // fix design matrix
    row_vector[NpredsV] V[Nobs]; // fix design matrix

    int<lower=1> Nlev1; // num center
    int<lower=1,upper=Nlev1> levind1[Nobs]; //center index
    int<lower=1,upper=N> levind2[Nobs];// patient index
    vector[Nobs] y; //continuous outcome
    int r[Nobs]; //binary outcome
    real<lower=0> sdscal; //overall residual

}

parameters {

    vector[NpredsX] alpha; //fixed LME coefs
    vector[NpredsV] beta; //fixed GLMM coefs
    real<lower=0> sigmalev1; // center sd
    real<lower=0> sigmalev2; // patient sd
    real<lower=0> sigmaeps; // epsilon sd
```

```
    real<lower=0> tau; // exp corr scale
    real<lower=0> rho; // exp corr 1/range
    vector[Nlev1] eta1; //latent center
    vector[N] eta2; //latent patient
    vector[Nobs] eta; //latent exp corr
    real<lower = -1, upper = 1> rho1; // assoc center param
    real<lower = -1, upper = 1> rho2; // assoc patient param
    real<lower=0> pow_r[Nlev1]; //power parameter

}

transformed parameters {

    vector[Nobs] w; // Gaussian process
    vector[Nobs] yhat; // linear predictor
    vector[Nlev1] ran1; // center effect
    vector[N] ran2; // patient effect
    vector[Nobs] Fsp; // response prob

    for (n in 1:N){

        matrix[T[n],T[n]] Sigma;
        matrix[T[n],T[n]] L_Sigma;
        vector[T[n]] t_sub;
        t_sub=t[start_pos[n]:(start_pos[n+1]-1)];

        //off-diagonal elements
        for(i in 1:(T[n]-1)){
             for (j in (i+1):T[n]){
                Sigma[i,j] = pow(tau,2) * exp(- rho * fabs(t_sub[i] -
                ↪  t_sub[j]));
                Sigma[j,i] = Sigma[i,j];
          }
        }

       // diagonal elements
        for (k in 1:T[n]){
                Sigma[k,k] = pow(tau,2)+0.000001; // + jitter
        }

        L_Sigma=cholesky_decompose(Sigma);
```

```
        w[start_pos[n]:(start_pos[n+1]-1)]= L_Sigma  *
        ↪  eta[start_pos[n]:(start_pos[n+1]-1)];
    }

    ran1 = sigmalev1 * eta1;
    ran2 = sigmalev2 * eta2;

    for(i in 1:Nobs){
        yhat[i] = X[i] * alpha + ran1[levind1[i]]  + ran2[levind2[i]] +
        ↪  w[i];
            Fsp[i]=spep_cdf(V[i] * beta + rho1 * ran1[levind1[i]] + rho2 *
            ↪  ran2[levind2[i]], pow_r[levind1[i]]);
    }
}

model {

    //----priors
    alpha ~ normal(0, 100);
    beta ~ normal(0, 100);
    eta1 ~ normal(0,1);
    eta2 ~ normal(0,1);
    eta ~ normal(0,1);
    sigmalev1 ~ cauchy(0, 5);
    sigmalev2 ~ cauchy(0, 5);
    sigmaeps ~ cauchy(0, 2.5*sdscal);
    rho1 ~ uniform(-1, 1);
    rho2 ~ uniform(-1, 1);
    pow_r ~ exponential(1);
    tau ~ normal(0, 5);
    rho ~ inv_gamma(2,1);

    //----likelihood

    y ~ normal(yhat,sigmaeps);
    r ~ bernoulli(Fsp);

}

generated quantities{
```

```
    vector[Nobs] log_lik_y;
    vector[Nobs] log_lik_r;

  for (i in 1:Nobs){

          log_lik_y[i] = normal_lpdf(y[i] | yhat[i], sigmaeps);
          log_lik_r[i] = bernoulli_lpmf(r[i] | Fsp[i]);

  }

}
```

B.6.2 R Code

```
library(rstan)
rstan::rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())

#read in STAN code
Stan.SPEP<-stan_model(stanc_ret=stanc("SPEP_JM4.stan"),verbose=FALSE)

FIT.JM4.SPEP<-sampling(Stan.SPEP,data=JD,warmup=2000,iter=5000,thin=3,chains=2,
↪  control=list(adapt_delta=0.95,max_treedepth=15),
              seed=1781002850, init_r=0.2,
              save_warmup=FALSE)
```

Appendix C

Appendix for Chapter 3

C.1 Conditional Survival Probability

$$
\begin{aligned}
S_{li}(t'|t) &= p(t_{n_{li}+1} \geq t' | t_{n_{li}+1} > t, \mathcal{D}) \qquad \text{(C.1)} \\
&= \int\int\int\int p(t_{n_{li}+1} \geq t' | t_{n_{li}+1} > t, b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta}, \mathcal{D}) \cdot p(b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta} | t_{n_{li}+1} > t, \mathcal{D})\, db_l\, d\boldsymbol{U}_{li}\, dv_{li}\, d\boldsymbol{\theta} \\
&= \int\int\int\int \frac{p(t_{n_{li}+1} \geq t' | b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta})}{p(t_{n_{li}+1} > t | b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta})} \cdot p(b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta} | t_{n_{li}+1} > t, \mathcal{D})\, db_l\, d\boldsymbol{U}_{li}\, dv_{li}\, d\boldsymbol{\theta} \\
&= \int\int\int\int \frac{S(t' | b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta})}{S(t | b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta})} \cdot p(b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta} | t_{n_{li}+1} > t, \mathcal{D})\, db_l\, d\boldsymbol{U}_{li}\, dv_{li}\, d\boldsymbol{\theta} \\
&= \int\int\int\int \frac{exp\Big[-\int_0^{t'} h(s | b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta}) ds \Big]}{exp\Big[-\int_0^{t} h(s | b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta}) ds \Big]} \cdot p(b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta} | t_{n_{li}+1} > t, \mathcal{D})\, db_l\, d\boldsymbol{U}_{li}\, dv_{li}\, d\boldsymbol{\theta} \\
&= \int\int\int\int exp\Big[-\int_t^{t'} h(s | b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta}) ds \Big] \cdot p(b_l, \boldsymbol{U}_{li}, v_{li}, \boldsymbol{\theta} | t_{n_{li}+1} > t, \mathcal{D})\, db_l\, d\boldsymbol{U}_{li}\, dv_{li}\, d\boldsymbol{\theta} \\
&\approx \frac{1}{M} \sum_{m=1}^{M} exp\Big[-\int_t^{t'} h\big(s | b_l^{(m)}, \boldsymbol{U}_{li}^{(m)}, v_{li}^{(m)}, \boldsymbol{\theta}^{(m)}\big) ds \Big]
\end{aligned}
$$

C.2 Simulation Result

Table C.1: Simulation results under models with current slope as association structure

Parameter	SLOPE+GAP						SLOPE+CALENDAR					
	Joint Model			Two-stage Method			Joint Model			Two-stage Method		
	Bias [a]	SE [b]	C90 [c]	Bias [a]	SE [b]	C90 [c]	Bias [a]	SE [b]	C90 [c]	Bias [a]	SE [b]	C90 [c]
Longitudinal submodel												
β_0	-0.20	0.36	0.86	-0.24	0.37	0.86	0.23	0.32	0.92	0.18	0.32	0.92
β_1	0.02	0.05	0.94	-0.11	0.06	0.92	0.09	0.06	0.88	0.10	0.07	0.84
β_2	-0.01	0.01	0.94	0.01	0.01	0.90	-0.01	0.01	0.88	0.00	0.01	0.88
σ	-0.19	0.06	0.90	-0.18	0.06	0.90	-0.07	0.06	0.90	-0.03	0.06	0.90
σ_b	-0.42	0.25	0.98	-0.39	0.24	0.98	-0.20	0.28	0.98	-0.21	0.28	0.96
σ_{u0}	0.06	0.11	0.94	0.09	0.11	0.94	-0.12	0.11	0.92	-0.14	0.12	0.92
σ_{u1}	0.04	0.02	0.92	0.03	0.02	0.90	0.01	0.02	0.96	-0.01	0.02	0.96
ρ	-0.01	0.01	0.90	-0.01	0.01	0.94	0.00	0.01	0.84	0.01	0.01	0.86
Event submodel												
γ_0	0.38	0.07	0.70	0.32	0.06	0.68	0.22	0.04	0.82	0.20	0.04	0.80
γ_1	0.01	0.04	0.88	0.00	0.04	0.88	-0.03	0.03	0.96	-0.03	0.03	0.96
γ_2	0.03	0.02	0.90	0.03	0.02	0.90	0.03	0.02	0.90	0.03	0.02	0.90
δ_1	-0.08	0.02	0.84	-0.07	0.02	0.84	-0.03	0.02	0.86	-0.02	0.02	0.88
δ_2	-0.05	0.02	0.90	-0.04	0.02	0.90	-0.01	0.02	0.90	-0.01	0.02	0.78
δ_3	-0.10	0.03	0.80	-0.07	0.03	0.82	-0.02	0.02	0.94	-0.02	0.02	0.94
δ_4	-0.08	0.03	0.84	-0.05	0.03	0.84	0.01	0.02	0.94	0.01	0.02	0.94
δ_5	-0.06	0.02	0.88	-0.05	0.02	0.90	0.00	0.03	0.90	0.01	0.03	0.88
δ_6	-0.07	0.02	0.88	-0.05	0.02	0.88	-0.04	0.03	0.88	-0.04	0.02	0.90
σ_v	-0.14	0.02	0.76	-0.15	0.02	0.78	-0.08	0.02	0.90	-0.10	0.02	0.92
Association structure												
α_1	0.02	0.01	0.88	0.01	0.01	0.84	0.01	0.01	0.98	0.01	0.01	0.94
α_2	0.01	0.01	0.88	0.00	0.01	0.88	0.02	0.01	0.94	0.02	0.01	0.92
α_3	-0.02	0.02	0.86	-0.01	0.02	0.86	0.02	0.02	0.86	0.02	0.02	0.86
α_4	-0.05	0.02	0.86	-0.04	0.02	0.84	0.00	0.02	0.94	0.00	0.02	0.92
α_5	-0.03	0.02	0.86	-0.03	0.02	0.88	0.01	0.01	0.94	0.01	0.01	0.96
α_6	-0.01	0.02	0.86	-0.01	0.02	0.90	-0.01	0.01	0.92	-0.01	0.01	0.92

[a] Bias=true value-mean estimate; [b] SE=standard error; [c] C90= probability of 90% coverage

Table C.2: Simulation results under models with current value as association structure

Parameter	VALUE+GAP						VALUE+CALENDAR					
	Joint Model			Two-stage Method			Joint Model			Two-stage Method		
	Bias [a]	SE [b]	C90 [c]	Bias [a]	SE [b]	C90 [c]	Bias [a]	SE [b]	C90 [c]	Bias [a]	SE [b]	C90 [c]
Longitudinal submodel												
β_0	0.07	0.33	0.90	0.08	0.33	0.92	-0.34	0.32	0.96	-0.30	0.32	0.96
β_1	0.05	0.06	0.96	0.20	0.06	0.90	0.00	0.05	0.92	0.11	0.06	0.92
β_2	0.00	0.01	0.86	-0.02	0.01	0.84	0.00	0.01	0.94	-0.02	0.01	0.90
σ	-0.07	0.05	0.90	-0.06	0.05	0.90	0.00	0.05	0.90	0.02	0.05	0.92
σ_b	-0.12	0.30	0.96	-0.18	0.28	0.96	-0.94	0.28	0.90	-0.89	0.28	0.92
σ_{u0}	-0.13	0.11	0.94	-0.15	0.11	0.94	0.00	0.11	0.92	-0.05	0.11	0.92
σ_{u1}	-0.01	0.02	0.98	-0.02	0.02	0.98	0.01	0.02	0.94	0.00	0.02	0.92
ρ	-0.01	0.01	0.94	-0.01	0.01	0.94	0.01	0.01	0.94	0.01	0.01	0.96
Event submodel												
γ_0	0.21	0.05	0.86	0.18	0.05	0.86	0.31	0.06	0.86	0.27	0.06	0.88
γ_1	-0.06	0.05	0.90	-0.06	0.05	0.90	-0.02	0.04	0.98	-0.02	0.04	0.98
γ_2	-0.03	0.02	0.96	-0.03	0.02	0.96	-0.04	0.02	0.92	-0.04	0.02	0.92
δ_1	-0.11	0.03	0.90	-0.10	0.03	0.86	-0.06	0.03	0.82	-0.05	0.03	0.80
δ_2	-0.03	0.03	0.90	-0.02	0.03	0.88	0.82	0.03	0.96	-0.07	0.03	0.98
δ_3	-0.07	0.02	0.90	-0.06	0.02	0.92	-0.04	0.03	0.92	-0.04	0.03	0.92
δ_4	-0.02	0.02	0.94	-0.02	0.02	0.94	-0.02	0.02	0.92	-0.02	0.02	0.92
δ_5	-0.06	0.02	0.88	-0.06	0.02	0.92	-0.05	0.02	0.96	-0.05	0.02	0.96
δ_6	-0.02	0.03	0.92	-0.01	0.03	0.90	-0.06	0.03	0.88	-0.05	0.03	0.86
σ_v	-0.16	0.03	0.80	-0.16	0.03	0.78	-0.05	0.03	0.88	-0.05	0.03	0.90
Association structure												
α_1	0.00	0.00	0.88	0.00	0.00	0.86	0.00	0.00	0.88	0.00	0.00	0.88
α_2	0.00	0.00	0.90	0.00	0.00	0.88	0.00	0.00	0.92	0.00	0.00	0.92
α_3	0.00	0.00	0.84	0.00	0.00	0.82	0.00	0.00	0.88	0.00	0.00	0.88
α_4	0.00	0.00	0.92	0.00	0.00	0.92	0.00	0.00	0.96	0.00	0.00	0.96
α_5	0.00	0.00	0.86	0.00	0.00	0.86	0.00	0.00	0.88	0.00	0.00	0.88
α_6	0.00	0.00	0.86	0.00	0.00	0.88	0.00	0.00	0.88	0.00	0.00	0.90

[a] Bias=true value-mean estimate; [b] SE=standard error; [c] C90=probability of 90% credible interval coverage

C.3 Data Cleaning

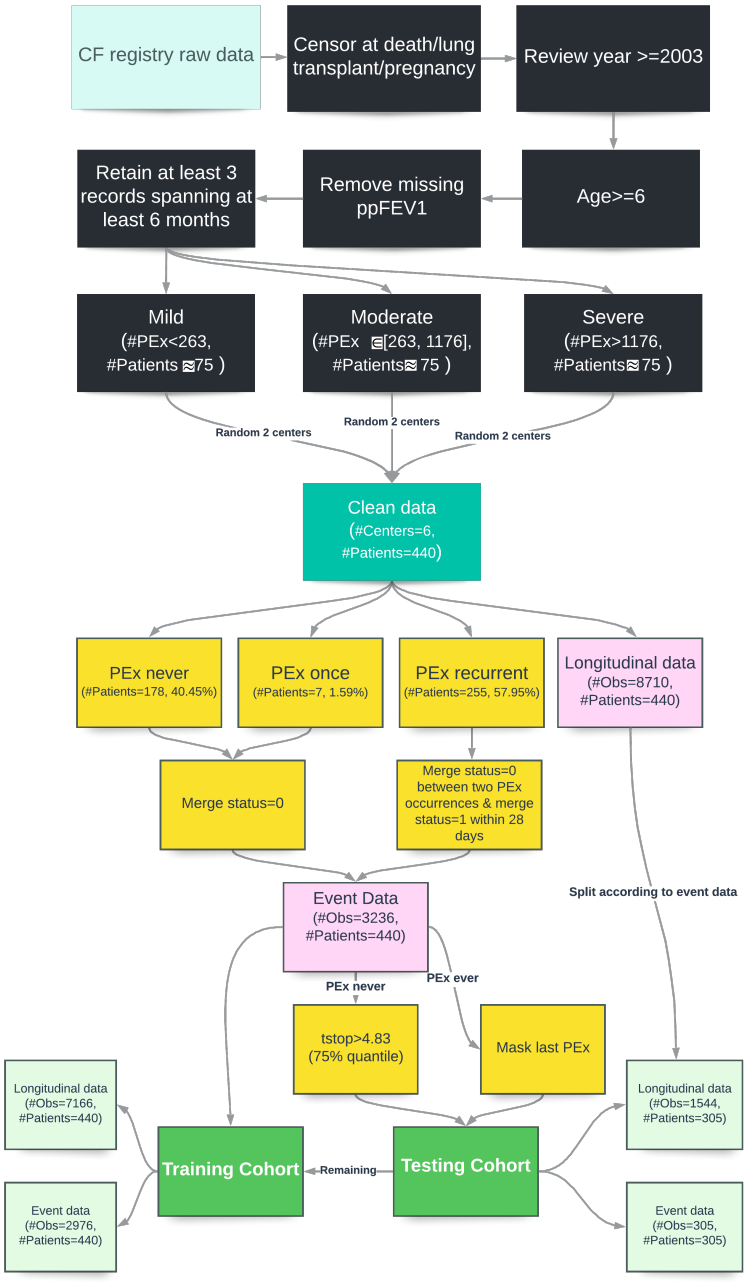

Figure C.1: Data cleaning process

C.4 Application Result

Table C.3: Estimation results under the model: SLOPE+GAP

Parameter	SLOPE+GAP					
	Joint Model			Two-stage Method		
	Mean [a]	SD [b]	CI [c]	Mean [a]	SD [b]	CI [c]
Longitudinal submodel						
β_0	7.20	1.66	[4.50, 9.84]	6.74	1.61	[4.19, 9.28]
β_1	0.85	0.02	[0.82, 0.88]	0.87	0.02	[0.84, 0.90]
β_2	-0.97	0.20	[-1.30, -0.64]	-1.15	0.23	[-1.52, -0.78]
β_3	-0.16	0.01	[-0.18, -0.14]	-0.10	0.02	[-0.13, -0.08]
β_4	-0.15	0.05	[-0.23, -0.08]	-0.20	0.04	[-0.27, -0.13]
β_5	3.32	1.13	[1.48, 5.13]	3.04	1.09	[1.22, 4.87]
β_6	3.83	1.19	[1.83, 5.70]	3.28	1.15	[1.44, 5.17]
β_7	5.18	1.21	[3.21, 7.17]	4.38	1.18	[2.44, 6.34]
β_8	-0.92	0.31	[-1.41, -0.42]	-0.91	0.32	[-1.43, -0.39]
β_9	1.75	0.83	[0.34, 3.07]	1.66	0.80	[0.41, 2.97]
σ	9.08	0.08	[8.95, 9.20]	9.05	0.08	[8.92, 9.18]
σ_b	1.96	1.12	[0.67, 4.06]	1.63	1.15	[0.23, 3.69]
σ_{u0}	6.89	0.31	[6.39, 7.41]	6.76	0.32	[6.24, 7.30]
σ_{u1}	2.94	0.17	[2.68, 3.22]	3.10	0.18	[2.82, 3.42]
ρ	0.16	0.07	[0.04, 0.27]	0.16	0.07	[0.03, 0.28]
Event submodel						
γ_0	0.31	0.18	[0.01, 0.61]	0.43	0.19	[0.12, 0.74]
γ_1	0.01	0.00	[0.01, 0.02]	0.02	0.00	[0.02, 0.03]
γ_2	0.41	0.24	[0.03, 0.81]	0.38	0.24	[-0.02, 0.78]
γ_3	-0.48	0.24	[-0.86, -0.08]	-0.53	0.23	[-0.91, -0.15]
δ_1	1.00	0.06	[0.91, 1.10]	1.01	0.05	[0.92, 1.09]
δ_2	1.23	0.07	[1.12, 1.35]	1.22	0.08	[1.10, 1.35]
δ_3	0.98	0.04	[0.92, 1.04]	0.96	0.04	[0.90, 1.02]
δ_4	1.06	0.04	[1.00, 1.12]	1.05	0.04	[1.00, 1.11]
δ_5	1.01	0.03	[0.96, 1.07]	1.01	0.03	[0.96, 1.07]
δ_6	0.88	0.03	[0.84, 0.93]	0.88	0.03	[0.84, 0.93]
σ_v	1.88	0.15	[1.65, 2.12]	2.06	0.15	[1.83, 2.32]
Association structure						
α_1	-0.34	0.09	[-0.50, -0.21]	-0.29	0.09	[-0.43, -0.15]
α_2	-0.52	0.13	[-0.74, -0.31]	-0.40	0.15	[-0.66, -0.15]
α_3	-0.47	0.06	[-0.58, -0.37]	-0.45	0.08	[-0.58, -0.33]
α_4	-0.47	0.07	[-0.59, -0.35]	-0.45	0.09	[-0.60, -0.31]
α_5	-0.43	0.07	[-0.55, -0.31]	-0.40	0.09	[-0.55, -0.26]
α_6	-0.28	0.05	[-0.37, -0.20]	-0.29	0.06	[-0.40, -0.18]

Note: all parameters converged at $\hat{R} \in [1.00, 1.02]$; [a] Mean=mean estimate; [b] SD=standard deviation; [c] CI= 90% credible interval

Table C.4: Estimation results under the model: VALUE+GAP

Parameter	VALUE+GAP					
	Joint Model			Two-stage Method		
	Mean [a]	SD [b]	CI [c]	Mean [a]	SD [b]	CI [c]
Longitudinal submodel						
β_0	7.62	1.54	[5.11, 10.18]	6.74	1.61	[4.19, 9.28]
β_1	0.86	0.02	[0.83, 0.89]	0.87	0.02	[0.84, 0.90]
β_2	-1.22	0.23	[-1.58, -0.84]	-1.15	0.23	[-1.52, -0.78]
β_3	-0.11	0.02	[-0.13, -0.08]	-0.10	0.02	[-0.13, -0.08]
β_4	-0.33	0.04	[-0.41, -0.26]	-0.20	0.04	[-0.27, -0.13]
β_5	3.17	1.08	[1.40, 4.97]	3.04	1.09	[1.22, 4.87]
β_6	3.65	1.16	[1.77, 5.59]	3.28	1.15	[1.44, 5.17]
β_7	4.68	1.13	[2.83, 6.53]	4.38	1.18	[2.44, 6.34]
β_8	-1.16	0.31	[-1.67, -0.66]	-0.91	0.32	[-1.43, -0.39]
β_9	1.56	0.82	[0.18, 2.88]	1.66	0.80	[0.41, 2.97]
σ	9.07	0.08	[8.94, 9.20]	9.05	0.08	[8.92, 9.18]
σ_b	1.67	1.05	[0.39, 3.56]	1.63	1.15	[0.23, 3.69]
σ_{u0}	6.76	0.32	[6.26, 7.30]	6.76	0.32	[6.24, 7.30]
σ_{u1}	3.01	0.17	[2.74, 3.30]	3.10	0.18	[2.82, 3.42]
ρ	0.17	0.07	[0.05, 0.29]	0.16	0.07	[0.03, 0.28]
Event submodel						
γ_0	4.17	0.22	[3.81, 4.54]	3.88	0.20	[3.58, 4.21]
γ_1	0.02	0.00	[0.02, 0.02]	0.02	0.00	[0.02, 0.03]
γ_2	0.25	0.20	[-0.08, 0.60]	0.26	0.19	[-0.04, 0.56]
γ_3	-0.34	0.18	[-0.62, -0.04]	-0.34	0.18	[-0.63, -0.05]
δ_1	1.01	0.06	[0.92, 1.11]	1.00	0.05	[0.92, 1.09]
δ_2	1.19	0.07	[1.08, 1.31]	1.17	0.07	[1.05, 1.29]
δ_3	1.01	0.04	[0.94, 1.07]	0.99	0.04	[0.93, 1.05]
δ_4	1.05	0.03	[0.99, 1.11]	1.04	0.03	[0.98, 1.10]
δ_5	1.01	0.03	[0.96, 1.06]	1.01	0.03	[0.95, 1.06]
δ_6	0.90	0.03	[0.86, 0.95]	0.89	0.03	[0.85, 0.94]
σ_v	1.59	0.11	[1.41, 1.76]	1.53	0.10	[1.36, 1.70]
Association structure						
α_1	-0.04	0.00	[-0.05, -0.03]	-0.04	0.00	[-0.04, -0.03]
α_2	-0.05	0.01	[-0.06, -0.05]	-0.05	0.01	[-0.06, -0.04]
α_3	-0.05	0.00	[-0.05, -0.04]	-0.04	0.00	[-0.05, -0.04]
α_4	-0.04	0.00	[-0.04, -0.03]	-0.03	0.00	[-0.04, -0.03]
α_5	-0.03	0.00	[-0.03, -0.02]	-0.02	0.00	[-0.03, -0.02]
α_6	-0.04	0.00	[-0.05, -0.04]	-0.04	0.00	[-0.04, -0.03]

Note: all parameters converged at $\hat{R} \in [1.00, 1.02]$; [a] Mean=mean estimate; [b] SD=standard deviation; [c] CI= 90% credible interval

Table C.5: Estimation results under the model: SLOPE+CALENDAR

Parameter	SLOPE+CALENDAR					
	Joint Model			Two-stage Method		
	Mean [a]	SD [b]	CI [c]	Mean [a]	SD [b]	CI [c]
Longitudinal submodel						
β_0	7.29	1.55	[4.77, 9.82]	6.74	1.61	[4.19, 9.28]
β_1	0.85	0.02	[0.83, 0.88]	0.87	0.02	[0.84, 0.90]
β_2	-1.17	0.22	[-1.53, -0.79]	-1.15	0.23	[-1.52, -0.78]
β_3	-0.13	0.02	[-0.16, -0.11]	-0.10	0.02	[-0.13, -0.08]
β_4	-0.16	0.05	[-0.23, -0.08]	-0.20	0.04	[-0.27, -0.13]
β_5	3.27	1.12	[1.47, 5.11]	3.04	1.09	[1.22, 4.87]
β_6	3.70	1.14	[1.82, 5.56]	3.28	1.15	[1.44, 5.17]
β_7	4.87	1.20	[2.89, 6.80]	4.38	1.18	[2.44, 6.34]
β_8	-0.92	0.31	[-1.44, -0.42]	-0.91	0.32	[-1.43, -0.39]
β_9	1.73	0.79	[0.44, 3.06]	1.66	0.80	[0.41, 2.97]
σ	9.06	0.08	[8.93, 9.20]	9.05	0.08	[8.92, 9.18]
σ_b	1.77	1.04	[0.54, 3.64]	1.63	1.15	[0.23, 3.69]
σ_{u0}	6.85	0.32	[6.35, 7.38]	6.76	0.32	[6.24, 7.30]
σ_{u1}	3.01	0.17	[2.75, 3.30]	3.10	0.18	[2.82, 3.42]
ρ	0.16	0.07	[0.06, 0.28]	0.16	0.07	[0.03, 0.28]
Event submodel						
γ_0	0.22	0.17	[-0.06, 0.51]	0.34	0.18	[0.04, 0.63]
γ_1	0.01	0.00	[0.01, 0.02]	0.01	0.00	[0.01, 0.02]
γ_2	0.46	0.23	[0.09, 0.84]	0.39	0.24	[0.01, 0.80]
γ_3	-0.53	0.22	[-0.89, -0.17]	-0.55	0.23	[-0.95, -0.18]
δ_1	1.29	0.09	[1.13, 1.44]	1.31	0.09	[1.16, 1.46]
δ_2	0.98	0.08	[0.85, 1.11]	0.99	0.08	[0.87, 1.12]
δ_3	1.40	0.10	[1.23, 1.56]	1.46	0.10	[1.31, 1.62]
δ_4	1.25	0.08	[1.13, 1.37]	1.28	0.08	[1.16, 1.40]
δ_5	1.19	0.07	[1.08, 1.30]	1.21	0.07	[1.11, 1.33]
δ_6	1.02	0.07	[0.90, 1.14]	1.04	0.07	[0.93, 1.15]
σ_v	1.86	0.13	[1.66, 2.07]	2.01	0.13	[1.80, 2.23]
Association structure						
α_1	-0.26	0.08	[-0.40, -0.13]	-0.20	0.08	[-0.34, -0.07]
α_2	-0.50	0.13	[-0.71, -0.29]	-0.43	0.16	[-0.70, -0.16]
α_3	-0.29	0.07	[-0.41, -0.18]	-0.23	0.08	[-0.37, -0.10]
α_4	-0.32	0.09	[-0.46, -0.18]	-0.27	0.09	[-0.43, -0.11]
α_5	-0.35	0.09	[-0.50, -0.21]	-0.30	0.10	[-0.46, -0.14]
α_6	-0.37	0.08	[-0.49, -0.24]	-0.37	0.08	[-0.51, -0.24]

Note: all parameters converged at $\hat{R} \in [1.00, 1.02]$; [a] Mean=mean estimate; [b] SD=standard deviation; [c] CI= 90% credible interval

Table C.6: Estimation results under the model: VALUE+CALENDAR

Parameter	VALUE+CALENDAR					
	Joint Model			Two-stage Method		
	Mean [a]	SD [b]	CI [c]	Mean [a]	SD [b]	CI [c]
Longitudinal submodel						
β_0	7.59	1.59	[5.09, 10.15]	6.74	1.61	[4.19, 9.28]
β_1	0.86	0.02	[0.83, 0.89]	0.87	0.02	[0.84, 0.90]
β_2	-1.12	0.22	[-1.48, -0.76]	-1.15	0.23	[-1.52, -0.78]
β_3	-0.11	0.02	[-0.14, -0.09]	-0.10	0.02	[-0.13, -0.08]
β_4	-0.33	0.05	[-0.40, -0.25]	-0.20	0.04	[-0.27, -0.13]
β_5	3.17	1.15	[1.28, 5.00]	3.04	1.09	[1.22, 4.87]
β_6	3.48	1.22	[1.32, 5.46]	3.28	1.15	[1.44, 5.17]
β_7	4.57	1.24	[2.49, 6.69]	4.38	1.18	[2.44, 6.34]
β_8	-1.15	0.30	[-1.64, -0.66]	-0.91	0.32	[-1.43, -0.39]
β_9	1.59	0.83	[0.23, 2.93]	1.66	0.80	[0.41, 2.97]
σ	9.07	0.08	[8.94, 9.20]	9.05	0.08	[8.92, 9.18]
σ_b	1.65	1.02	[0.37, 3.59]	1.63	1.15	[0.23, 3.69]
σ_{u0}	6.78	0.31	[6.27, 7.29]	6.76	0.32	[6.24, 7.30]
σ_{u1}	3.02	0.18	[2.73, 3.33]	3.10	0.18	[2.82, 3.42]
ρ	0.18	0.07	[0.06, 0.30]	0.16	0.07	[0.03, 0.28]
Event submodel						
γ_0	3.66	0.23	[3.28, 4.06]	3.46	0.22	[3.11, 3.82]
γ_1	0.01	0.00	[0.00, 0.02]	0.01	0.00	[0.01, 0.02]
γ_2	0.29	0.19	[-0.02, 0.60]	0.29	0.19	[-0.01, 0.60]
γ_3	-0.33	0.19	[-0.64, -0.02]	-0.38	0.18	[-0.67, -0.08]
δ_1	1.24	0.09	[1.10, 1.38]	1.25	0.09	[1.11, 1.40]
δ_2	1.01	0.08	[0.88, 1.13]	1.01	0.08	[0.88, 1.13]
δ_3	1.28	0.07	[1.16, 1.40]	1.30	0.07	[1.18, 1.42]
δ_4	1.26	0.06	[1.16, 1.36]	1.26	0.06	[1.16, 1.36]
δ_5	1.24	0.06	[1.15, 1.34]	1.24	0.06	[1.15, 1.34]
δ_6	1.07	0.05	[0.98, 1.15]	1.07	0.05	[0.98, 1.16]
σ_v	1.60	0.10	[1.44, 1.77]	1.56	0.10	[1.41, 1.74]
Association structure						
α_1	-0.04	0.00	[-0.05, -0.03]	-0.03	0.00	[-0.04, -0.03]
α_2	-0.05	0.01	[-0.06, -0.04]	-0.04	0.01	[-0.05, -0.04]
α_3	-0.04	0.00	[-0.05, -0.04]	-0.04	0.00	[-0.05, -0.03]
α_4	-0.04	0.00	[-0.04, -0.03]	-0.03	0.00	[-0.04, -0.03]
α_5	-0.03	0.00	[-0.03, -0.02]	-0.02	0.00	[-0.03, -0.02]
α_6	-0.04	0.00	[-0.05, -0.04]	-0.04	0.00	[-0.04, -0.03]

Note: all parameters converged at $\hat{R} \in [1.00, 1.02]$; [a] Mean=mean estimate; [b] SD=standard deviation; [c] CI= 90% credible interval

C.5 Convergence Diagnostics

In this section, we have investigated some common visual MCMC diagnostics using the R bayesplot (Gabry and Mahr (2020)) package for our optimal model. The time series plot of the Markov chains is shown in Figure C.2. Typically we can see that both chains explore the similar region of parameter values, which is a good sign. We can also visualize the ACF for each Markov chain separately up to a lag of 20 for each parameter. We prefer ACF to drop quickly to zero with increasing lag because positive autocorrelation means the chain tends to stay in the same area between iterations. All parameters are shown to meet this expectation from Figure C.3 and Figure C.4. It is notable that negative autocorrelation is possible and it indicates fast convergence of sample mean towards true mean.

The motivation of the potential scale reduction statistic($\hat{R}$) is to measure the ratio of the average variance of draws within each chain to the variance of pooled draws across chains. If the chains have not converged to a common distribution, the $\hat{R}$ statistic will be greater than one (Gelman et al. (2013b),Stan Development Team (2020)). The points in Figure C.5 representing $\hat{R}$ values are colored based on some cutoffs and there are no divergences observed.

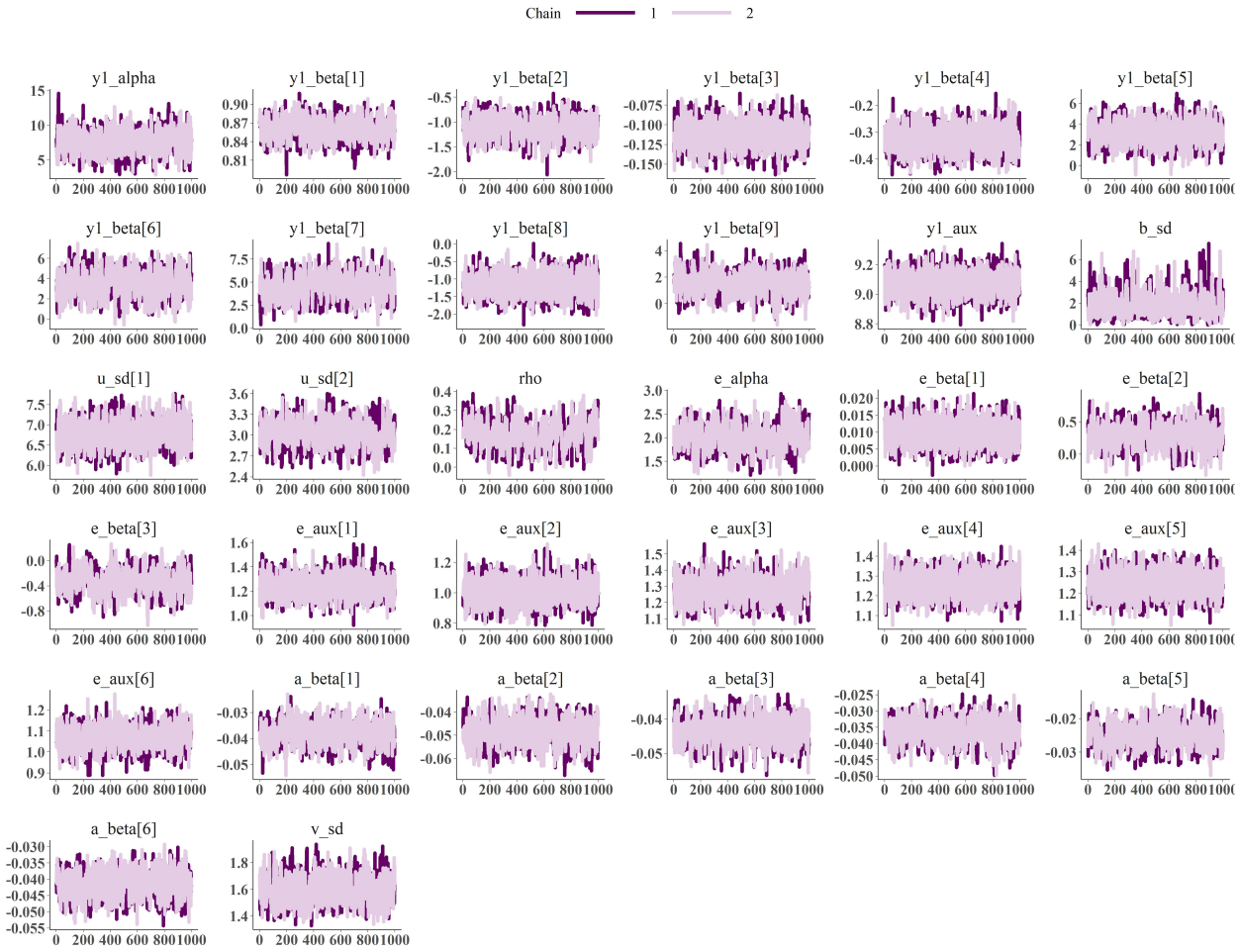

Figure C.2: Traceplot against iterations

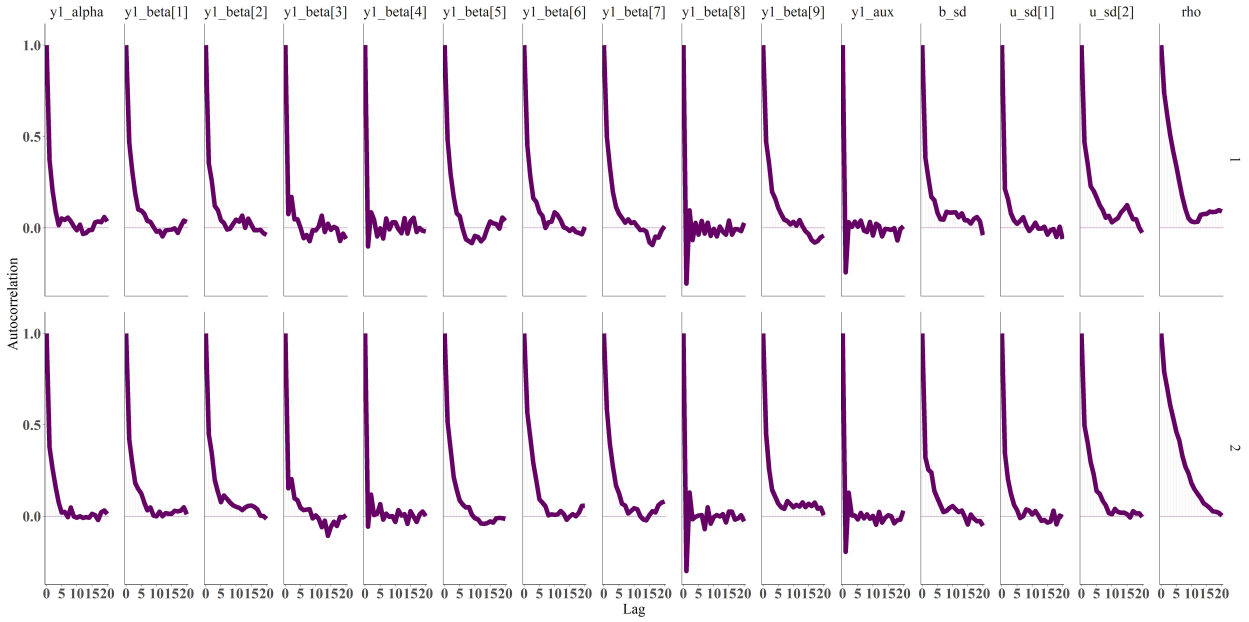

Figure C.3: Autocorrelation for parameters from longitudinal submodel

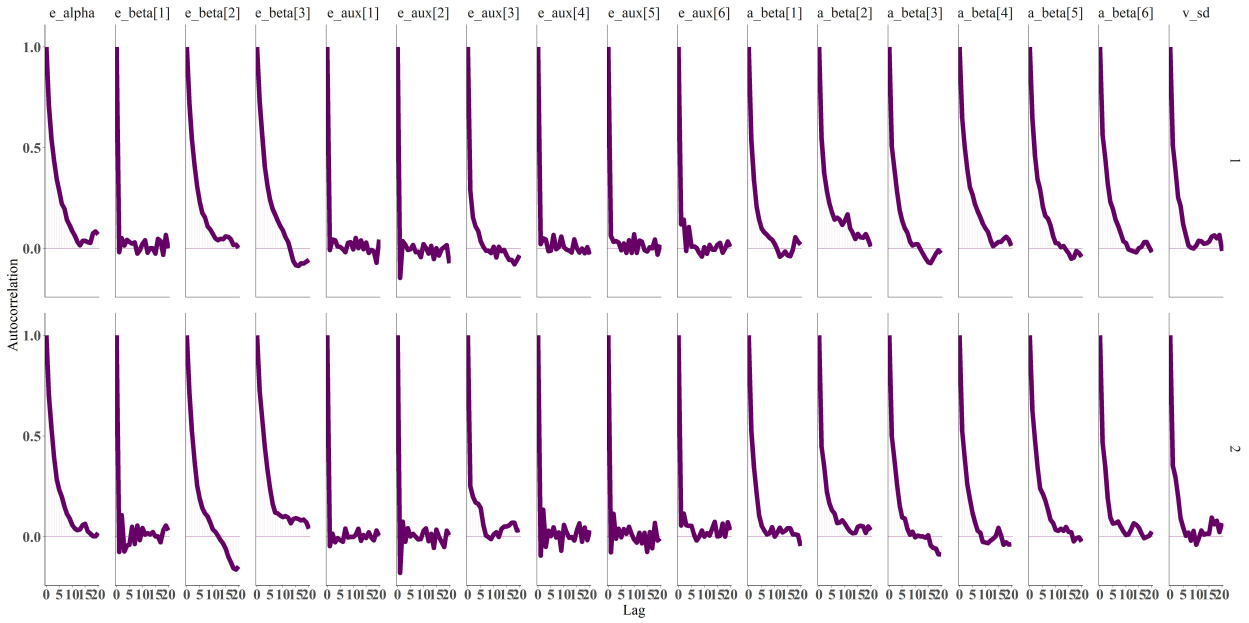

Figure C.4: Autocorrelation for parameters from event submodel

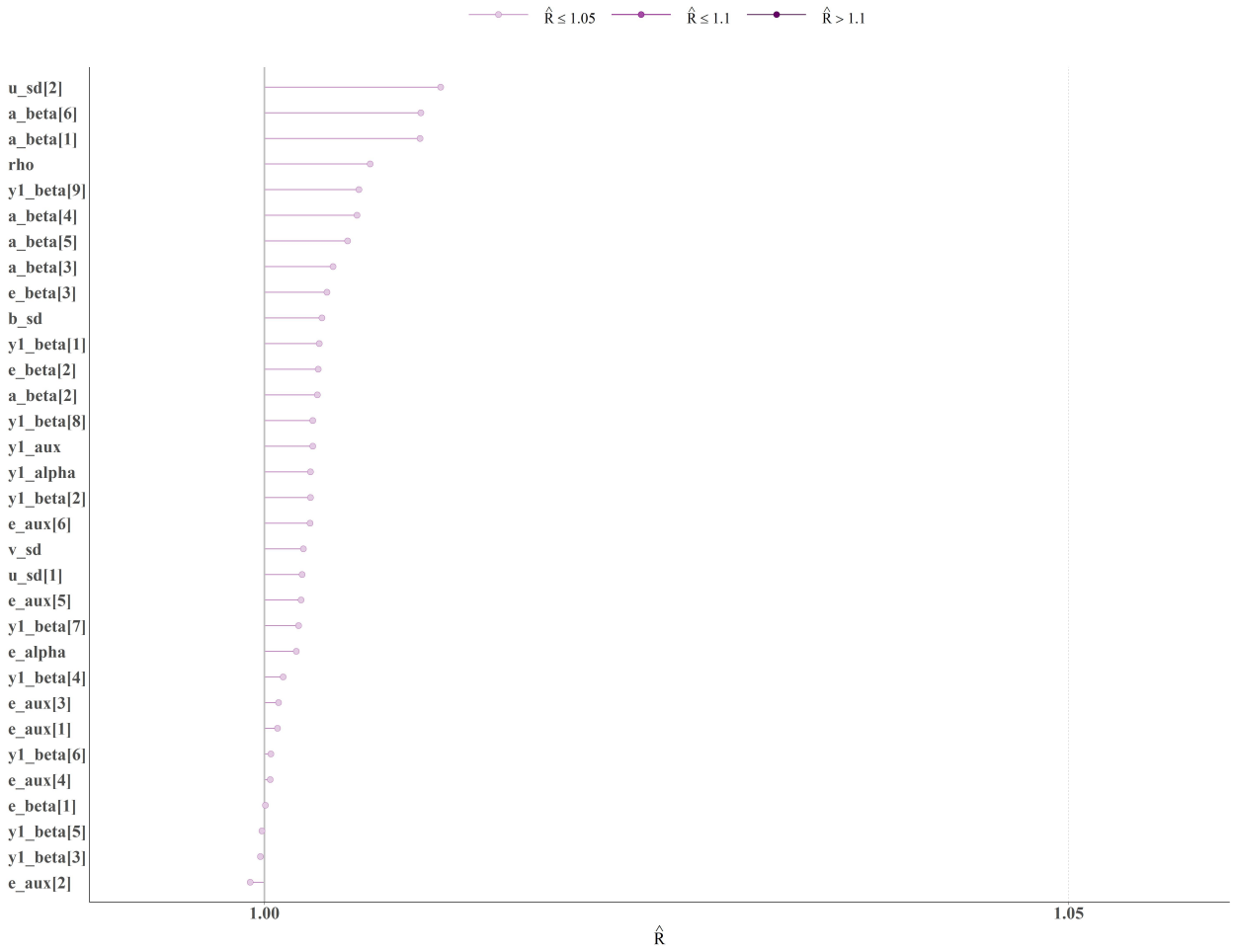

Figure C.5: Rhat plot

C.6 Time and System

Additional information about processing system and time are described in Table C.7 and Table C.8. All models are estimated with 4000 iterations through two chains. The first 2000 draws are discarded as a warm-up sampling and the remaining 2000 are kept for the posterior inference for both simulated data and real data.

Table C.7: Processing system and versions

	Simulated data	**Real data**
Platform	x86_64-apple-darwin17.0 (64-bit)	x86_64-w64-mingw32/x64 (64-bit)
Running under	macOS Big Sur 10.16	Windows 10 x64 (build 19043)
R version	4.0.5 (2021-03-31)	4.0.2 (2020-06-22)
CmdStan	v2.28.2	v2.29.1
cmdstanr	v0.4.0	v0.5.0

Table C.8: Elapsed time for different models

Association + Time scale	Model	Simulated data (hrs/rep) (N = 480)	Real data (hrs) (N = 440)
Slope + Gap	Joint Model	0.18	9.16
	Two-stage Method	0.05	2.86
Slope + Calendar	Joint Model	0.26	3.82
	Two-stage Method	0.08	2.59
Value + Gap	Joint Model	0.18	8.17
	Two-stage Method	0.04	5.31
Value + Calendar	Joint Model	0.15	7.10
	Two-stage Method	0.04	4.90

hrs=hours; rep=replicate; N=Number patient

C.7 Example Code

We attach an example code for illustration purpose. For the complete code package, please refer to my Github.

C.7.1 Stan Program

```
/************************************************/
// Purpose: Fit proposed JM in the manuscript
// Association structure: Slope/Value
// Risk interval: Gap/Calendar
// Assumption: i) full conditional independence;
//             ii) LME: random intercept-slope;
```

```
//             iii) Survival: extended stratified relative risk frailty
↪  model
// Author: Copyright (C) 2022 Grace C. Zhou
// Date: Mar., 2022
// Based upon:
// Copyright (C) 2015, 2016, 2017 Trustees of Columbia University
// Copyright (C) 2016, 2017 Sam Brilleman

/************************************************/

functions {

  vector evaluate_eta(matrix X, array[] vector Z_u, int Dev_index,

                      array[] int U_id, array[] int C_id, real gamma,

                      vector beta, vector bVec, matrix uMat) {

    int N = rows(X); // num rows in design matrix

    int K = rows(beta); // num predictors

    //int p = size(Z_u);    // num group level params:intercept+slope

    vector[N] eta;

    if (K > 0) {

      eta = X * beta + gamma * Dev_index;

    } else {

      eta = rep_vector(0.0, N) + gamma * Dev_index;

    }

    //for (k in 1:p)

    for (n in 1 : N) {

      eta[n] = eta[n] + bVec[C_id[n]] * Dev_index
```

```
            + uMat[U_id[n], 1] * Z_u[1, n] + uMat[U_id[n], 2] * Z_u[2,
            ↪  n];
  }

  return eta;
}

/**
* Get the indices corresponding to the lower tri of a square matrix
* @param dim The number of rows in the square matrix
* @return A vector of indices
*/
array[] int lower_tri_indices(int dim) {
  array[dim + choose(dim, 2)] int indices;
  int mark = 1;
  for (r in 1 : dim) {
    for (c in r : dim) {
      indices[mark] = (r - 1) * dim + c;
      mark = mark + 1;
    }
  }
  return indices;
```

```
  }
}
data {
  //----- Longitudinal submodels
  // population level dimensions
  int<lower=0> y_N; // num observations
  int<lower=0> y_K; // num predictors

  // population level data
  vector[y_N] y1; // response vectors
  matrix[y_N, y_K] y1_X; // fix effect design matrix
  vector[y_K] y1_Xbar; // predictor means

  // group level dimensions
  int<lower=0> b_N; // num center
  int<lower=0> b_K; // num center predictor
  int<lower=0> u_N; // num patients
  int<lower=0> u_K;// num patient predictor
  array[y_N] int<lower=0> y1_C_id; // center id
  array[y_N] int<lower=0> y1_U_id; // patient id
  array[u_K] vector[y_N] y1_Z; // random effect design matrix
```

```
//----- Event submodel

// data for calculating event submodel linear predictor in GK quadrature

// NB these design matrices are evaluated AT the event time and

// the (unstandardised) quadrature points

int<lower=0> e_K; // num predictors

int<lower=0> a_K; // num assoc params

int<lower=0> Npat; // num patients (equal to u_N)

int<lower=0> Nevents; // num events (ie. not censored)

int<lower=0> qnodes; // num nodes for GK quadrature

int<lower=0> Nobs_times_qnodes; // Nobs x qnodes

int<lower=0> nrow_e_Xq; // num rows predictor matrix

matrix[nrow_e_Xq, e_K] e_Xq; // design matrix

vector[e_K] e_Xbar; // predictor means

real norm_const; // norm constant

int<lower=0> basehaz_df; // baseline hazard df

vector[nrow_e_Xq] basehaz_X; // baseline hazard design matrix/vector
↪  (basis terms)

vector[Nobs_times_qnodes] qwts; // GK quadrature weights with (b-a)/2
↪  scaling

matrix[nrow_e_Xq, y_K] y1_Xq; // fix effect design matrix at quadpoints

array[u_K] vector[nrow_e_Xq] y1_Zq; // random effect design matrix at
↪  quadpoints
```

```
  array[nrow_e_Xq] int<lower=0> y1_Cq_id; // center id at quadpoints
  array[nrow_e_Xq] int<lower=0> y1_Uq_id; // patient id at quadpoints
  //----- Hyperparameters for prior distributions
  // scale prior
  vector<lower=0>[y_K] y1_prior_scale;
  vector<lower=0>[e_K] e_prior_scale;
  vector<lower=0>[a_K] a_prior_scale;
  real<lower=0> y_prior_scale_for_intercept;
  real<lower=0> e_prior_scale_for_intercept;
  real<lower=0> y_prior_scale_for_aux;
  vector<lower=0>[basehaz_df] e_prior_scale_for_aux;

  // lkj prior
  real<lower=0> b_prior_scale;
  vector<lower=0>[u_K] u_prior_scale;
  vector<lower=0>[u_K] u_prior_df;
  real<lower=0> u_prior_regularization;
  int<lower=0> Dev_index;
}
transformed data {
  // indexing used to extract lower tri of RE covariance matrix
```

```
  array[u_K + choose(u_K, 2)] int u_cov_idx;
  if (u_K > 0) {
    u_cov_idx = lower_tri_indices(u_K);
  }
}
parameters {
 //----- Longitudinal submodel
  real y1_gamma; // intercept
  vector[y_K] y1_z_beta; // unscaled coef
  real<lower=0> y1_aux_unscaled; // unscaled residual error
  real<lower=0> b_sd; // center sd
  vector[b_N] z_b_vec; // unscaled center effect
  matrix[u_K, u_N] z_u_mat; // unscaled patient effect
  vector<lower=0>[u_K] u_sd; // patient sd
  cholesky_factor_corr[u_K > 1 ? u_K : 0] u_cholesky; // cholesky factor
 //----- Event submodel
  real e_gamma; // intercept in event submodel
  vector[e_K] e_z_beta; // unscaled log hazard ratio
  vector<lower=0>[basehaz_df] e_aux_unscaled; // unscaled baseline hazard
  ↪  coef
  vector[a_K] a_z_beta; // unscaled assoc params
```

```
  real<lower=0> v_sd; // frailty sd
  vector[Npat] z_v_vec;// unscaled frailty effect
}
transformed parameters {
 //----- Longitudinal submodel
  vector[y_K] y1_beta; // scaled coef
  real<lower=0> y1_aux; // scaled residual error
  matrix[u_N, u_K] u_mat; // patient effect
  vector[b_N] b_vec; // scaled center effect
  vector[y_N] y1_eta; // linear predictor
 //----- Event submodel
  vector[nrow_e_Xq] y1_eta_q; // linear predictor at quadpoints
  vector[e_K] e_beta; // scaled coef (log hazard ratio)
  vector[a_K] a_beta; // scaled assoc params
  vector<lower=0>[basehaz_df] e_aux; // scaled baseline hazard coef
  vector[Npat] v_vec; // scaled frailty effect
  //----- Longitudinal submodel
  y1_beta = y1_z_beta .* y1_prior_scale;
  y1_aux = y1_aux_unscaled * y_prior_scale_for_aux;
  b_vec = b_sd * z_b_vec;
  if (u_K == 1) {
```

```
    u_mat = (u_sd[1] * z_u_mat)';
  } else if (u_K > 1) {
    u_mat = (diag_pre_multiply(u_sd, u_cholesky) * z_u_mat)';
  }
  //----- Event submodel
  e_beta = e_z_beta .* e_prior_scale;
  a_beta = a_z_beta .* a_prior_scale;
  e_aux = e_aux_unscaled .* e_prior_scale_for_aux;
  v_vec = v_sd * z_v_vec;
  //----- Longitudinal submodel
  // linear predictor at observed time
  y1_eta = evaluate_eta(y1_X, y1_Z, 1, y1_U_id, y1_C_id, y1_gamma, y1_beta,
                        b_vec, u_mat);
 //----- Event submodel
 // linear biomarker predictor at event time and quadrature points
  y1_eta_q = evaluate_eta(y1_Xq, y1_Zq, Dev_index, y1_Uq_id, y1_Cq_id,
                          y1_gamma, y1_beta, b_vec, u_mat);
}
model {
  //---- Longitudinal submodel
```

```
 // increment the target with the log-lik

 target += normal_lpdf(y1 | y1_eta, y1_aux);

//---- Event submodel (Gauss-Kronrod quadrature)

 {

   vector[nrow_e_Xq] e_eta_q; // linear predictor at event time and
   ↪  quadrature points

   vector[nrow_e_Xq] log_basehaz; // log baseline hazard at event time and
   ↪  quadrature points

   vector[nrow_e_Xq] log_haz_q; // log hazard at event time and quadrature
   ↪  points

   vector[Nevents] log_haz_etimes; // log hazard at the event time only

   vector[Nobs_times_qnodes] log_haz_qtimes; // log hazard at the
   ↪  quadrature points only

   for (n in 1 : nrow_e_Xq) {

   // Step 1: event submodel add on contribution from association structure
   ↪  to
   // the linear predictor at event time and quadrature points

     e_eta_q[n] = e_Xq[n] * e_beta + y1_eta_q[n] * a_beta[y1_Cq_id[n]]

                  + v_vec[y1_Uq_id[n]];

     // Step 2: log baseline hazard (Weibull) at event time and quadrature
     ↪  points

     log_basehaz[n] = e_gamma + norm_const + log(e_aux[y1_Cq_id[n]])

                      + basehaz_X[n] * (e_aux[y1_Cq_id[n]] - 1);

   }
```

```
    // Step 3: log hazard at event time and quadrature points
    log_haz_q = log_basehaz + e_eta_q;
    // Step 4: log hazard at event times only
    log_haz_etimes = head(log_haz_q, Nevents);
    // Step 5: log hazard at quadrature points only
    log_haz_qtimes = tail(log_haz_q, Nobs_times_qnodes);
    // Step 6: increment the target with the log-lik
    target += sum(log_haz_etimes) - dot_product(qwts, exp(log_haz_qtimes));
  }
  //---- Log-priors
 //----- Longitudinal submodel
  target += normal_lpdf(y1_gamma | 0, y_prior_scale_for_intercept);
  target += normal_lpdf(y1_z_beta | 0, 1);
  target += normal_lpdf(y1_aux_unscaled | 0, 1);
  target += normal_lpdf(z_b_vec | 0, 1);
  target += normal_lpdf(b_sd | 0, b_prior_scale); // following Gelman 2008
  target += student_t_lpdf(u_sd | u_prior_df, 0, u_prior_scale);
  target += normal_lpdf(to_vector(z_u_mat) | 0, 1);
  // corr matrix
  if (u_K > 1) {
    target += lkj_corr_cholesky_lpdf(u_cholesky | u_prior_regularization);
```

```
  }
 //----- Event submodel
  target += normal_lpdf(e_gamma | 0, e_prior_scale_for_intercept);
  target += normal_lpdf(e_z_beta | 0, 1);
  target += normal_lpdf(a_z_beta | 0, 1);
  target += normal_lpdf(e_aux_unscaled | 0, 1);
  target += normal_lpdf(z_v_vec | 0, 1);
  target += normal_lpdf(v_sd | 0, 10);
}

generated quantities {
  real y1_alpha; // transformed intercept for long submodel
  vector[size(u_cov_idx)] u_cov; // var-cov for patient
  real rho; // correlation coef
  real e_alpha; // transformed intercept for event submodel
  //---- Long submodel
  y1_alpha = y1_gamma - dot_product(y1_Xbar, y1_beta);
  // Transform variance-covariance matrix for patient
  if (u_K == 1) {
    u_cov[1] = u_sd[1] * u_sd[1];
  } else {
```

```
    u_cov = to_vector(quad_form_diag(multiply_lower_tri_self_transpose(
                                        u_cholesky), u_sd))[u_cov_idx];
  }

  rho = u_cov[2] / (u_sd[1] * u_sd[2]);
   for (i in 1 : y_N) {
    log_lik_y[i] = normal_lpdf(y1[i] | y1_eta[i], y1_aux); // log-lik
    y_tilde[i] = normal_rng(y1_eta[i], y1_aux); // posterior predictive dist
  }
 //---- Event submodel
  e_alpha = e_gamma + norm_const - dot_product(e_Xbar, e_beta);
}
```

C.7.2 R Code

```
library(cmdstanr)

file.jm <- file.path("JM.stan")
mod.jm <- cmdstan_model(file.jm)

fit.jm <- mod.jm$sample(
  data = standata.jm,
  chains = 2,
  save_warmup = FALSE,
  parallel_chains = 2,
  refresh = 500,
  adapt_delta=0.95,
  max_treedepth=12,
  seed=202207,
```

```
  init = function() staninit.jm
)
```