Erasmus University Rotterdam

Master Quantitative Marketing and Business Analytics

Econometric Institute

Personalized screening intervals for biomarkers: a joint-modeling application for heart failure patients

Author:
Tobias Polak

Supervisor
Prof. Dr. R. Paap
Co-reader
Prof. Dr. D. Fok

Supervisors MC:
Prof. Dr. Ir. H. Boersma
Dr. D. Rizopoulos
Dr. J. Van Rosmalen

Abstract

Screening for diseases is common practice in illness detection. The design of optimal, personalized screening intervals has received more attention as personalized medicine has become more popular. In this work, we focus on the modeling of longitudinal biomarker measurements. We extend the framework of joint modeling in the field of screening intervals of Rizopoulos et al. (2016) in two directions. First, we consider a Bayesian model average specification. Second, we allow for the simultaneous scheduling of multiple screenings. We illustrate the use of our adaptions with an application among heart failure patients and the NT-proBNP biomarker. We find that (i) higher levels of the biomarker places the patient at greater risk for cardiac events and (ii) that Bayesian model averaging allows for modeling non-standard biomarker trajectories.

Keywords: Personalized medicine, Bayesian model averaging, Random effects.

August 20, 2016

Contents

1 Introduction **2**

2 Conceptual framework **3**

3 Joint modeling framework **5**
3.1 Longitudinal Analysis 6
3.1.1 Estimation of the mixed model 6
3.2 Survival Analysis 7
3.2.1 Parameter estimation of the survival process 8
3.3 Joint model 9
3.3.1 Maximum likelihood estimation 9
3.3.2 Bayesian estimation 10

4 Personalized screening intervals **11**
4.1 Model selection 11
4.2 Scheduling of the next measurement 12
4.3 Estimation of the Kullback-Leibner divergence 14

5 Extensions **15**
5.1 Bayesian model averaging 16
5.2 Scheduling multiple screenings 17

6 Results **19**
6.1 Comparison of cvDCL and BMA 20
6.2 Illustration of two upcoming measurements 25

7 Discussion **26**

8 Conclusion **27**

A Appendix I **31**
A.1 Predicted survival probabilities 31
A.2 Scheduling S screenings 31

B Appendix II **33**
B.1 Results of patient 19 33
B.2 EKL$(s_1, s_2 \mid t)$ for patient 23 34

1 Introduction

As detection of illnesses in time increases the life expectancy of patients through earlier treatment, routinely screening for diseases such as breast cancer among older asymptomatic women is nowadays common practice. The same holds for patients known to have had prostate cancer: monitoring the prostate-specific antigen provides information on the current prostate condition. Scheduling these screenings is part of the field of medical decision making. This field has become more and more complex, due to an increasing amount of data available from patients all over the world and the development of different treatment options. As the Internet has become overgrown with uncensored medical information, today's patient has become more involved in his/her medical process. This has led to shared decision making, where physicians and patients together assess the desired treatment. Along with the popularization and effectiveness of tailor-made treatments, or personalized medicine, (bio-)statisticians also play an important role in medical decision making. The analysis of patient data leads to new optimal and personalized medical screening strategies.

The statistical analysis of screening strategies and optimization thereof has been subject of research in the last 20 years. Parmigiani (1993, 1998, 2002) and Lee and Zelen (1998) use dynamical optimization and multi-state Markov-models. These models can handle altering number of screenings to plan as well as different treatment options. The optimal screening strategy provides the most information on disease progression at acceptable costs: the chosen model is most cost-effective. Of course, cost is measured not only in monetary terms but also in quality of life from a patients perspective. Recently, the joint modeling of longitudinal and survival data (e.g. Tsiatis and Davidian (2004) and Rizopoulos (2012)) has been used to optimally schedule an upcoming screening one-step ahead (Rizopoulos et al., 2016). As joint models include a random effects specification (Laird and Ware, 1982), the nature of these models is subject-specific. Modeling of individual patient trajectories can help in obtaining personalized screening intervals whereas usually, screening procedures are homogeneous over patients, with fixed time intervals. Moreover, joint models take the whole history of measurements into account, whilst many screening strategies are merely based on the latest available measurement. Finally, Rizopoulos et al. (2016) use information theory quantities to define an optimal model, bearing in mind the patient's condition.

In this thesis, we will extend the work of Rizopoulos et al. (2014) and Rizopoulos et al. (2016) in two directions. Firstly, instead of focusing on the best predictive model as Rizopoulos et al. (2016), we combine different model specifications via Bayesian model averaging, following Hoeting et al. (1999) and Rizopoulos et al. (2014). Thereupon, we decide a new screening time. In this way, we do not limit ourselves in model choice. Secondly, the framework of Rizopoulos et al. (2016) allows only for the planning of one upcoming measurement. We will adapt their framework to fit simultaneous planning of multiple screenings at once.

The bioSHiFT studies of the Department of Cardio-Thoracic Surgery of Erasmus Medican Center (EMC) among 263 patients form the motivation of our thesis. These

patients are diagnosed with possible coronary artery disease and/or chronic heart failure. Coronary artery disease (CAD) is an inflammatory disease of the endothelium of the coronary arteries (Mueller, 2014). The inflammation may lead to ruptures of vulnerable lesions and tissues. These ruptures induce intracoronary thrombus formation and acute coronary syndrome (ACS), including myocardial infarction. Chronic heart failure (CHF) is characterized by a malfunctioning cardiac muscle, resulting in a shortage of blood flow throughout the body (Mueller, 2014). Heart failure (HF) is often the common result of the majority of heart diseases, as different underlying processes can lead to pump failure. In 2012 in the Netherlands, CAD/ACS were responsible for 6.000 deaths and 84.000 hospital admissions. HF was responsible for 6.800 deaths and 29.000 hospital admissions (Hartstichting, 2014). Repeated measures of blood and urines samples yield information on biomarker levels that could possibly indicate a primary endpoint: re-operation, hospital admission or death. Biomarkers, or biochemical markers, that we consider in this setting are (levels of) plasma proteins that reflect pathopysiological mechanisms. Routinely, patients are screened for biomarkers to determine different treatment options and health risks. As an earlier mentioned example, prostate-specific antigen (PSA) levels are indicative of the status of prostate cancer in treated men. For CAD/CHF patients, cardiac troponin-T (cTnT) and the N-terminal pro B-type natriuretic peptide (NT-proBNP) are relevant biomarkers on which clinical guidelines have based patient management programmes (Libby, 2006). In this study, we limit ourselves to investigate the relation between cardiac events and NT-proBNP, and schedule multiple novel screenings.

The remainder of this thesis is organized as follows. Section 2 will give a conceptual overview of the methods used. In Section 3, we will discuss the full joint model and its derivation. Section 4 will focus on the theory of personalized screening intervals. Section 5 will discuss our extensions, the results of which are depicted in Section 6. In Section 7, we will shed a light on the idea and definition of optimality in the field of scheduling screenings. Finally, Section 8 concludes.

2 Conceptual framework

Measurements of biomarkers and changes thereof provide physicians insight in the health conditions of patients. Therefore, patients have to plan multiple visits to the hospital to give blood or plasma samples. Biomarker levels fluctuate over these samples. This fluctuation is partly due to between-subject differences such as age or gender, but the levels of biomarkers usually do not remain constant within a patient either. Furthermore, not every patient will have the same risks associated to similar biomarker values. We will try to model the subject-specific trajectory of NT-proBNP and relate levels of the biomarker to health risks. Health risks are measured in terms of an event. An event is defined as death, operation and/or hospital admission. Figure 1 depicts NT-proBNP levels of the event-group and the event-free group. Clearly, we observe a higher average level of NT-proBNP in the event-group. We will make use of a mixed effects specification (see Laird and Ware (1982) or Verbeke and Molenberghs (2000) for details) to model

the longitudinal trajectory, in combination with a hazard-specification (Fleming and Harrington, 1991) to model the risks of events occurring. The simultaneous estimation of both results in the joint model. For details, we refer to Rizopoulos (2012).

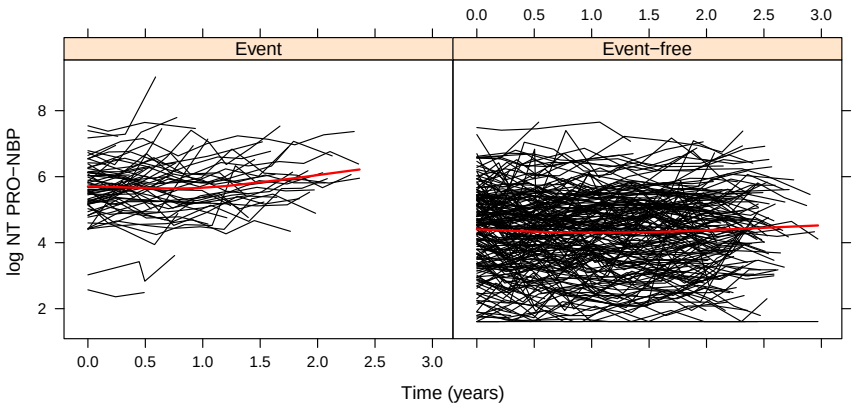

Figure 1: The trajectory of NT-proBNP in the event-group (left) and non-event group (right). The red line shows a fitted average trajectory.

The joint model links a longitudinal path (biomarker measurements) to a risk-describing process: hazard. The intuitive idea is depicted in Figure 2. Of course, a variety of different relations between the longitudinal outcome and the hazard function exists. We will consider multiple joint model specifications. Afterwards, we will pick the optimal model(s) for each patient at every visit. Optimality will be assessed on predictive power or standard Bayesian selection criteria, such as deviance information criteria (DIC) or log pseudo-marginal likelihood (LPML). Finally, multiple models will be combined in a Bayesian averaging framework.

The planning of screenings is crucial for patients. On the one hand, over-frequent hospital visits are a burden for both patients and physicians, and health care costs rise. On the other hand, when patients are barely examined, they suffer great risks of cardiac events. Clearly, this involves a trade-off between burden/costs and risks. Another aspect that has to be taken into account when planning screening intervals, is the information gained by a new measurement. Measuring a patient multiple times per day will yield low health risks, but due to low intra-day variation between these measurements, phsyicians do not gain much insight into the global shape of the longitudinal trajectory. Ideally, we would plan the next measurement such that we gain as much insight as possible, whilst bearing in mind the health risks of every patient. Given the biomarker trajectory at time t, we would like to plan a measurement within $t + \Delta t$ whilst ensuring that at $t + \Delta t$, the chances of survival are still above a certain threshold κ. Figure 3 elucidates the setting.

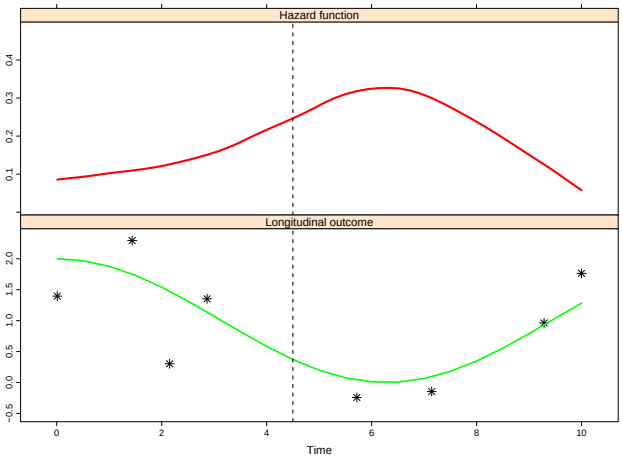

Figure 2: Schematic exemplary overview of the joint model. The top panel displays the path of the hazard function. The bottom panel shows the longitudinal trajectory of a biomarker.

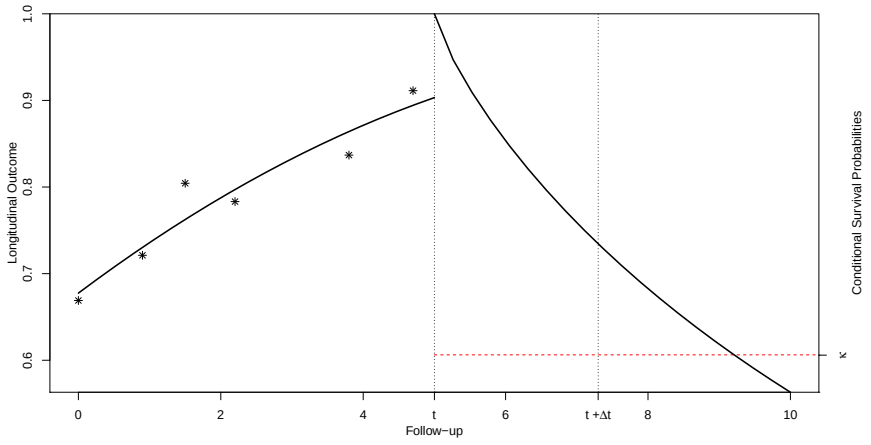

Figure 3: Planning of a new measurement at timepoint $t + \Delta t$. The left-hand panel displays the longitudinal measurements of a biomarker, the solid line on the right-hand side displays the survival function. The dashed red line depicts the threshold-survival value.

3 Joint modeling framework

This section consists of three parts. In the first part, we will discuss the modeling of the longitudinal trajectory. In the second part, we will elaborate on survival analysis.

Based on the first two parts, we will introduce the resulting the joint model. We will consider both maximum likelihood as Bayesian estimation techniques.

3.1 Longitudinal Analysis

The linear mixed-effects model is a standard framework for longitudinal data analysis, see Verbeke and Molenberghs (2000). We denote $\mathbf{y}_i$ the $n_i \times 1$ vector of follow-up measurements for subject $i, i = 1, \ldots, N$. y_{il} denotes the outcome at timepoint $t_{il}, l = 1, ..., n_i$. The linear mixed model entails:

$$\begin{aligned} y_i(t) &= m_i(t) + \varepsilon_i(t) & \varepsilon_i(t) &\sim \mathcal{N}(0, \sigma^2) \\ &= \boldsymbol{x}_i(t)^\intercal \boldsymbol{\beta} + \boldsymbol{z}_i(t)^\intercal \boldsymbol{u}_i + \varepsilon_i(t) & \boldsymbol{u}_i &\sim \mathcal{N}(\mathbf{0}, \mathbf{D}) \end{aligned} \tag{3.1}$$

where in (3.1) $m_i(t)$ is the *true* underlying value of the longitudinal outcome measured with a normally distributed error $\varepsilon_i(t)$ with mean 0 and variance σ^2. $\boldsymbol{x}_i(t) \in \mathbb{R}^{K \times 1}$ and $\boldsymbol{z}_i(t) \in \mathbb{R}^{Q \times 1}$ denote (time-varying) design matrices associated with the fixed effects $\boldsymbol{\beta}$ and random effects $\boldsymbol{u}_i$, respectively. The random effects follow a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{D}$ and are independent of both the error terms $\varepsilon_i(t)$ and the fixed effects. Implicitly, (3.1) accounts for correlation between measurements within the same subject through the random effects. An advantage of the mixed-effects model is that it allows for estimation of parameters for the population mean via the fixed effects, and additionally models the individual trajectory over time via the random effects.

An important feature of the mixed model is that the outcomes are marginally correlated through the random effects, but they are assumed to be independent *given* the random effects:

$$p(\boldsymbol{y}_i \mid \boldsymbol{u}_i) = \prod_{l=1}^{n_i} p(y_{il} \mid \boldsymbol{u}_i) \tag{3.2}$$

3.1.1 Estimation of the mixed model

Maximum likelihood (ML) estimators of the parameters above are most frequently used. The marginal likelihood contribution of the ith subject is given by

$$p(\boldsymbol{y}_i) = \int_{\boldsymbol{u}_i} p(\boldsymbol{y}_i \mid \boldsymbol{u}_i) p(\boldsymbol{u}_i) \, d\boldsymbol{u}_i. \tag{3.3}$$

and the above has a closed-form solution, leading to

$$\boldsymbol{y}_i \sim \mathcal{N}(\boldsymbol{X}_i^\intercal \boldsymbol{\beta}, \underbrace{\boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^\intercal + \sigma^2 I_{n_i}}_{\boldsymbol{V}_i}) \tag{3.4}$$

and hence the log-likelihood contribution of subject i equals

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \sum_{i=1}^{N} \log\, p(\boldsymbol{y}_i \mid \boldsymbol{\theta}) \\ &= \sum_{i=1}^{N} \log\, \left[(2\pi)^{-\frac{n_i}{2}} \left|\boldsymbol{V}_i\right|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\intercal \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\}\right] \end{aligned} \tag{3.5}$$

The corresponding maximum likelihood estimator of $\boldsymbol{\beta}$ is equal to the generalized least squares estimator:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i^\intercal \boldsymbol{V}_i^{-1} \boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i^\intercal \boldsymbol{V}_i^{-1} \boldsymbol{y}_i \tag{3.6}$$

when $\boldsymbol{V}_i$ is known. When only estimates $\hat{\boldsymbol{V}}_i$ are available, the estimated variance-covariance matrix can replace the true in (3.6), asymptotically yielding unbiased estimates. In small samples, one often opts for restricted maximum likelihood (REML) estimates (Harville, 1974). In this way, one obtains unbiased estimates of the variance/covariance components. The log-likelihood optimized via REML estimation is:

$$\begin{aligned}\mathcal{L}_{RE}(\boldsymbol{\theta}) &= -\frac{N-K}{2}\log(2\pi) + \frac{1}{2}\log\left|\sum_{i=1}^{N} \boldsymbol{X}_i^\intercal \boldsymbol{X}_i\right| - \frac{1}{2}\log\left|\sum_{i=1}^{N} \boldsymbol{X}_i^\intercal \boldsymbol{V}_i \boldsymbol{X}_i\right| \\ &\qquad - \frac{1}{2}\sum_{i=1}^{N}\{\log\left|\boldsymbol{V}_i\right| + (\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}})^\intercal \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}})\} \\ &\propto -\frac{1}{2}\log\sum_{i=1}^{N}|\boldsymbol{V}_i| - \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}})^\intercal \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) \\ &\qquad - \frac{1}{2}\log\left|\sum_{i=1}^{N} \boldsymbol{X}_i^\intercal \boldsymbol{V}_i^{-1} \boldsymbol{X}_i\right| \end{aligned} \tag{3.7}$$

where in (3.7) $\hat{\boldsymbol{\beta}}$ is given by (3.6). Optimization of the latter will give estimates of $\boldsymbol{V}_i$ that account for the fact that $\boldsymbol{\beta}$ was estimated in (3.6). The restricted likelihood resembles concentrated/profile likelihood approaches, but penalizes by the last term of (3.7).

3.2 Survival Analysis

For the modeling of the time to re-operation/death/hospital admission, we make use of basic concepts of survival analysis. For a detailed overview, we refer to Fleming and Harrington (1991). We denote T_i^* the true event time of subject i, C_i the censoring time and $T_i = \min(T_i^*, C_i)$ the observed event time. Hence, the observed event time

equals true event time as long as the subject is still in our study. We denote the survival function

$$S_i(t) = \Pr(T_i^* > t) = \int_t^\infty f(s)\ ds \tag{3.8}$$

where $f(s)$ is the probability density function of the event. The hazard function describes the risk of an instantaneous event in the time frame Δt, given that subject i is event-free up to time t:

$$\lambda_i(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T_i^* < t + \Delta t \mid T_i^* \geq t)}{\Delta t} = \frac{f(t)}{S(t)}. \tag{3.9}$$

By definition, the survival function (3.8) can be expressed by means of its corresponding hazard (3.9) via:

$$S_i(t) = \exp\left\{ - \int_0^t \lambda_i(s)\ ds \right\} \tag{3.10}$$

where the bracketed term is known as the cumulative hazard function. Different specifications for the hazard exist. In this thesis, we will make use of a proportional hazard function:

$$\lambda_i(t) = \lambda_0(t)\ \exp\{\boldsymbol{\gamma}^\intercal \boldsymbol{w}_i\} \tag{3.11}$$

where $\lambda_0(t)$ denotes the *baseline* hazard, $\boldsymbol{w}_i$ denotes a vector of explanatory variables and $\boldsymbol{\gamma}$ the associated parameters.

3.2.1 Parameter estimation of the survival process

Parameter estimation of the survival process is usually done via maximum likelihood. For notation purposes, let the spell-indicator $\delta_i = I(T_i^* \leq C_i)$ indicate whether the event has occurred or not. Here, $I(A)$ is the logical operator, equal to 1 when A is true, 0 else wise. The log-likelihood is

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \log \prod_{i=1}^{N} p(T_i, d_i \mid \boldsymbol{\theta}) \\
&= \sum_{i=1}^{N} \log f(T_i \mid \boldsymbol{\theta})^{\delta_i} S(T_i \mid \boldsymbol{\theta})^{1-\delta_i} \\
&= \sum_{i=1}^{N} \delta_i \log \lambda_i(T_i \mid \boldsymbol{\theta}) - \int_0^{T_i} \lambda_i(s \mid \boldsymbol{\theta}) ds.
\end{aligned} \tag{3.12}$$

3.3 Joint model

The joint model relates the longitudinal outcome to the survival process. By simultaneously modeling the longitudinal outcome and incorporating it in the time-to-event model, one is able to measure the association between both outcomes. Let $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s \leq t\}$ be the history of the true longitudinal process. We denote the set of all available information $\mathcal{D}_n = \{\boldsymbol{y}_i, T_i, \delta_i; i = 1, \ldots, N\}$. We will relate the survival sub-model to the longitudinal outcome via the hazard function:

$$\lambda_i(t \mid \mathcal{M}_i(t), \boldsymbol{w}_i) = \lambda_0(t)\exp\{\boldsymbol{\gamma}^\intercal \boldsymbol{w}_i + f(m_i(t), \boldsymbol{u}_i, \boldsymbol{\alpha})\} \tag{3.13}$$

where in (3.13) $\boldsymbol{\alpha}$ measures the association between the (functional form of the) longitudinal process $m_i(t)$, the random effects $\boldsymbol{u}_i$ and the hazard. Following Rizopoulos (2012), we consider different relations between the longitudinal process and the survival outcome:

$$\begin{aligned} &f(m_i(t), \boldsymbol{u}_i, \alpha) = \alpha m_i(t), && f(m_i(t), \boldsymbol{u}_i, \alpha) = \alpha m_i'(t) \\ &f(m_i(t), \boldsymbol{u}_i, \alpha) = \alpha \int_0^t m_i(s)\ ds, \end{aligned} \tag{3.14}$$

where $m_i'(t) = \frac{\partial m_i(t)}{\partial t}$ is the direction of the marker. The above relations either depend on *(i)* the underlying value at point (t), *(ii)* the derivative of the longitudinal outcome, or *(iii)* the accumulated survival process up to t. Of course, combinations of the above can also be used. Note that the proportional lifetime model (3.14) *(i)* states that the instantaneous risk of an event only depends upon the value of the process at time t. This is not the case for the survival function, since

$$\begin{aligned} S_i(t \mid \mathcal{M}_i(t), \boldsymbol{w}_i) &= p(T_i^* > t \mid \mathcal{M}_i(t), \boldsymbol{w}_i) \\ &= \exp\left(-\int_0^t \lambda_0(s)\exp\{\boldsymbol{\gamma}^\intercal \boldsymbol{w}_i + f(m_i(s), \boldsymbol{u}_i, \boldsymbol{\alpha})\}ds\right) \end{aligned} \tag{3.15}$$

implies that the event-to-time process is related to the entire trajectory $\mathcal{M}_i(t)$. As $m_i(t)$ is a time-varying variable, we need the integral of (3.15) to be proper. This is the case as long as the time-varying process is bounded and predictable, for the proof see Fleming and Harrington (1991), page 131. For the full joint distribution that describes both the longitudinal and time-to-event processes, we will need one extra assumption next to (3.2). Similarly for the survival process, we will assume that the random effects account for all correlation between the longitudinal and survival outcomes. That is, the event process and the longitudinal trajectory are assumed to be independent given $\boldsymbol{u}_i$:

$$p(T_i, \delta_i, \boldsymbol{y}_i \mid \boldsymbol{u}_i; \boldsymbol{\theta}) = p(T_i, \delta_i \mid \boldsymbol{u}_i; \boldsymbol{\theta})p(\boldsymbol{y}_i \mid \boldsymbol{u}_i, \boldsymbol{\theta}) \tag{3.16}$$

3.3.1 Maximum likelihood estimation

Previous research has proposed different estimation methods. A two-stage method based on first estimating the longitudinal outcome via (3.7) and then using these estimates to

obtain second-stage estimates of (3.13) does not take into account the uncertainty in the first-stage regression. A full, joint-likelihood approach has been followed by Rizopoulos (2011), Rizopoulos et al. (2009) and Faucett and Thomas (1996), among others. Under the assumptions (3.2) and (3.16), we note the ith contribution to the full log-likelihood

$$\begin{aligned}\log p(T_i, \delta_i, \boldsymbol{y}_i \mid \boldsymbol{\theta}) &= \log \int_{\boldsymbol{u}_i} p(T_i, \delta_i, \boldsymbol{y}_i \mid \boldsymbol{\theta}) d\boldsymbol{u}_i \\ &= \log \int_{\boldsymbol{u}_i} \underbrace{p(T_i, \delta_i \mid \boldsymbol{u}_i, \boldsymbol{\theta})}_{\text{survival}} \underbrace{\left[\prod_{l=1}^{n_i} p(y_{il} \mid \boldsymbol{u}_i, \boldsymbol{\theta})\right] p(\boldsymbol{u}_i \mid \boldsymbol{\theta})}_{\text{longitudinal}} d\boldsymbol{u}_i. \end{aligned} \tag{3.17}$$

The survival part of (3.17) follows from (3.13) and (3.15):

$$p(T_i, \delta_i \mid \boldsymbol{u}_i, \boldsymbol{\theta}) = \lambda_i(T_i \mid \mathcal{M}_i(T_i), \boldsymbol{\theta})^{\delta_i} S_i(T_i \mid \mathcal{M}_i(T_i), \boldsymbol{\theta}), \tag{3.18}$$

which is similar to the likelihood contribution of (3.12). For the longitudinal response, we observe

$$\begin{aligned} p(\boldsymbol{y}_i \mid \boldsymbol{u}_i, \boldsymbol{\theta}) p(\boldsymbol{u}_i \mid \boldsymbol{\theta}) &= \prod_{l=1}^{n_i} p(y_{il} \mid \boldsymbol{u}_i, \boldsymbol{\theta}) p(\boldsymbol{u}_i \mid \boldsymbol{\theta}) \\ &= (2\pi\sigma^2)^{-\frac{n_i}{2}} \exp\{-\frac{1}{2}\sigma^2 ||\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{u}_i||\} \\ &\quad \times (2\pi)^{-\frac{q}{2}} |\boldsymbol{D}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}\boldsymbol{u}_i^\intercal \boldsymbol{D}^{-1}\boldsymbol{u}_i\}. \end{aligned} \tag{3.19}$$

The maximization of (3.17) requires the computation of numerous integrals. Since we need to integrate out all random effects, the numerical computation can become a burden. Therefore, we resort to Bayesian estimation techniques to overcome this problem.

3.3.2 Bayesian estimation

For a Bayesian approach, we observe that the full posterior is of the form:

$$p(\boldsymbol{\theta}, \boldsymbol{u}_i \mid \mathcal{D}_n) \propto \prod_{i=1}^{N} \prod_{l=1}^{n_i} p(y_{il} \mid \boldsymbol{u}_i, \boldsymbol{\theta}) p(T_i, \delta_i \mid \boldsymbol{u}_i, \boldsymbol{\theta}) p(\boldsymbol{u}_i \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}). \tag{3.20}$$

Clearly, we still need to specify the exact form of the baseline hazard. Following Rizopoulos et al. (2016), we opt for a flexible spline specification. For details on splines, we refer to Eilers and Marx (1996). The full conditional likelihood contribution of subject i is of

the form:

$$
\begin{aligned}
p(\boldsymbol{y}_i, T_i, \delta_i \mid \boldsymbol{u}_i, \boldsymbol{\theta}) &= \prod_{l=1}^{n_i} p(y_{il} \mid \boldsymbol{u}_i, \boldsymbol{\theta}) p(T_i, \delta_i \mid \boldsymbol{u}_i, \boldsymbol{\theta}) p(\boldsymbol{u}_i \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \\
&\propto (2\pi\sigma^2)^{-\frac{n_i}{2}} \exp\{-\frac{1}{2}\sigma^2 ||\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{u}_i||\} \\
&\times (2\pi)^{-\frac{q}{2}} |\boldsymbol{D}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}\boldsymbol{u}_i^\intercal \boldsymbol{D}^{-1}\boldsymbol{u}_i\} \\
&\times \left[\exp\left\{\sum_q B_q(T_i, \boldsymbol{v}) + \boldsymbol{\gamma}^\intercal \boldsymbol{w}_i + f(m_i(T_i), \boldsymbol{u}_i, \boldsymbol{\alpha})\right\}\right]^{\delta_i} \\
&\times \exp\left[-\exp(\boldsymbol{\gamma}^\intercal \boldsymbol{w}_i) \int_0^{T_i} \exp\left\{\sum_q B_q(s, \boldsymbol{v}) + \boldsymbol{\gamma}^\intercal \boldsymbol{w}_i + f(m_i(s), \boldsymbol{u}_i, \boldsymbol{\alpha})\right\} ds\right] \\
&\times p(\boldsymbol{\theta}). \qquad (3.21)
\end{aligned}
$$

The prior distribution of parameters $\boldsymbol{\theta}$ is left unspecified above. We specify diffuse proper priors for all regression coefficients, parametrization is discussed in the Results Section. The conditional posterior distribution of all terms including $\boldsymbol{u}_i$ are not of a known form. Realizations are obtained via a random walk Metropolis-Hastings sampler, and tuned via scaling the variance-covariance matrix $\boldsymbol{V}_i$ of (3.4). For the (cumulative) hazard, $B_q(T_i, v)$ denotes the q-the basis of a B-spline. Using a flexible spline specification for the baseline hazard is common practice, see Rizopoulos (2012) or Eilers and Marx (1996).

4 Personalized screening intervals

This section will be dedicated to the planning of a new measurement within the framework of joint models. More precisely, how do we plan the next measurement of patient j, given his previous biomarker levels. First, the appropriate model is selected for patient j. The models will differ in association structure, as in (3.14). Thereafter, we determine a quantity that yields the best next measurement time. Notation is based on the work of Rizopoulos et al. (2016).

4.1 Model selection

To choose the correct personalized screening interval for the individual subject j at time t, one has to choose an appropriate model. The joint model entails both specification of the mixed and survival model. Standard Bayesian evaluation is based on deviance information criterion (DIC) or Bayes-factors, taking into account both the fit for the longitudinal as time-to-event outcomes. For a personalized screening interval however, one might be inclined to favor the model that predicts future events best, given that subject j is event-free up to time t. Let $\mathcal{M} = \{M_1, \ldots, M_k\}$ be the set of possible models, and M^* the true data generating model. To address the predictive power for the survival outcome per model, we make use of the cross-validatory posterior predictive

conditional density of the time-to-event process. For the kth model, this is equal to $p(T_i^* \mid T_i^* > t, \mathcal{Y}_i(t), \mathcal{D}_{n-i}, M_k)$ where $\mathcal{D}_{n-i}$ denotes the full dataset when subject i is left out and $\mathcal{Y}_i(t) = \{y_i(t_{il}); 0 \leq t_{il} \leq t, l = 1, \ldots, n_i\}$ the history of longitudinal measurements for subject i. Rizopoulos et al. (2016) choose the model M_k that minimizes the cross-entropy (Cover and Thomas, 1991)

$$\mathrm{CE}_k(t) = \mathbb{E}\left\{-\log\left[p(T_i^* \mid T_i^* > t, y_i(t), \mathcal{D}_{n-i}, M_k)\right]\right\}. \tag{4.1}$$

The expectation is taken with respect to the predictive conditional density under the *true* model, i.e. $p(T_i^* \mid T_i^* > t, \mathcal{Y}_i(t), \mathcal{D}_{n-i}, M^*)$. Rizopoulos et al. (2016) estimate (4.1) by the *cross-validated Dynamic Conditional Likelihood*:

$$\mathrm{cvDCL}_k(t) = \frac{1}{n_t} \sum_{i=1}^{N} -I(T_i > t) \log p(T_i, \delta_i \mid T_i > t, \mathcal{Y}_i(t), \mathcal{D}_{n-i}, M_k), \tag{4.2}$$

whilst taking the absence of subject i and possible censoring into account. In (4.2), n_t equals the number of subjects that have survived up to time t, $n_t = \sum_i I(T_i > t)$. An MCMC-sample of the equation above can be drawn by use of importance sampling:

$$\begin{aligned}
[p(T_i, \delta_i \mid T_i > t, \mathcal{Y}_i(t), \mathcal{D}_{n-i})]^{-1} &= \frac{p(\mathcal{D}_{n-i}, T_i > t, \mathcal{Y}_i(t))}{p(T_i, \delta_i, T_i > t, \mathcal{Y}_i(t), \mathcal{D}_{n-i})} \\
&= \int_{\boldsymbol{\theta}} \frac{p(\mathcal{D}_{n-i} \mid \boldsymbol{\theta}) p(T_i > t, \mathcal{Y}_i(t) \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathcal{D}_n)} d\boldsymbol{\theta} \\
&= \int_{\boldsymbol{\theta}} \frac{p(T_i > t, \mathcal{Y}_i(t) \mid \boldsymbol{\theta})}{p(T_i, \delta_i, \mathcal{Y}_i(t) \mid \boldsymbol{\theta})} \frac{p(\mathcal{D}_n) \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathcal{D}_n)} d\boldsymbol{\theta} \\
&= \int_{\boldsymbol{\theta}} \frac{1}{p(T_i, \delta_i \mid T_i > t, \mathcal{Y}_i(t), \boldsymbol{\theta})} p(\boldsymbol{\theta} \mid \mathcal{D}_n) d\boldsymbol{\theta}
\end{aligned} \tag{4.3}$$

And hence Monte Carlo estimates of (4.2) can be found via:

$$\widehat{\mathrm{cvDCL}}(t) = \frac{1}{n_t} \sum_{i=1}^{N} I(T_i > t) \log \left(\frac{1}{G} \sum_{g=1}^{G} \frac{1}{p(T_i, \delta_i \mid T_i > t, y_i(t), \boldsymbol{\theta}^{(g)})} \right), \tag{4.4}$$

where $\boldsymbol{\theta}^{(g)}$ denotes the gth draw of the posterior distribution $p(\boldsymbol{\theta} \mid \mathcal{D}_n)$, derived from (3.20). Finally, we choose the model that maximizes the estimate (4.4). Note that this is a harmonic mean estimator (Newton and Raftery, 1994) for which the central limit theorem need not hold. The variances of the weights need to be checked for stability of the estimator. For a detailed reference on limit conditions, see Wolpert and Schmidler (2011).

4.2 Scheduling of the next measurement

For the scheduling of the next measurement for subject j, we will rely upon the observed data of this patient $\{T_j^* > t, \mathcal{Y}_j(t)\}$ combined with the model selected in the previous

Section for this patient at time point t . The scheduling of the next measurement at timepoint $u, u > t$ involves a risk/gain-tradeoff. On the one hand, we would like to maximize the information gain obtained by a new measurement $y_j(u)$, given that the patient did not experience the event up to timepoint u. On the other hand, if the event actually occurs before u, there will be no information gain. Therefore, we also take into account the risk of event occurrence. Rizopoulos et al. (2016) make use of concepts in Bayesian optimal designs (Clyde and Chaloner, 1996) to model the risk/gain tradeoff in choosing the optimal u. The utility function considered equals

$$U(u \mid t) = \mathbb{E}\left\{ \lambda_1 \underbrace{\log \frac{p(T_j^* \mid T_j^* > u, \{\mathcal{Y}_j(t), y_j(u)\}, \mathcal{D}_n)}{p(T_j^* \mid T_j^* > u, \mathcal{Y}_j(t), \mathcal{D}_n)}}_{\text{Information ratio}} - \lambda_2 \underbrace{I(T_j^* > u)}_{\text{Risk term}} \right\}, \tag{4.5}$$

where the first information ratio is equal to the expected Kullback-Leibner divergence, i.e. the difference between the posterior predictive conditional distribution with a measurement at timepoint u and without. As we condition on the measurement $y_j(u)$, it makes sense to condition on $T_j^* > u$, as measurements are only possible as long as the patient is alive. Formally, the Kullback-Leibner divergence equals

$$\begin{aligned} \text{EKL}(u \mid t) &= \mathbb{E}_Y \left[\mathbb{E}_{T^*|Y} \left\{ \log \frac{p(T_j^* \mid T_j^* > u, \{\mathcal{Y}_j(t), y_j(u)\}, \mathcal{D}_n)}{p(T_j^* \mid T_j^* > u, \mathcal{Y}_j(t), \mathcal{D}_n)} \right\} \right] \\ &= \int_{y_j(u)} \left\{ \int_{T_j^*} \log \frac{p(T_j^* \mid T_j^* > u, \{\mathcal{Y}_j(t), y_j(u)\}, \mathcal{D}_n)}{p(T_j^* \mid T_j^* > u, \mathcal{Y}_j(t), \mathcal{D}_n)} p(T_j^* \mid T_j^* > t, \{\mathcal{Y}_j(t), y_j(u)\}, \mathcal{D}_n) dT_j^* \right\} \\ &\qquad \times p(y_j(u) \mid T_j^* > t, \mathcal{Y}_j(t), \mathcal{D}_n) dy_j(u). \end{aligned} \tag{4.6}$$

An increase in this ratio adheres to an increase in information gain. If the event takes place before u, then $p(T_j^* \mid T_j^* > u, \{\mathcal{Y}_j(t), y_j(u)\}, \mathcal{D}_n)$ simplifies to $p(T_j^* \mid T_j^* > u, \mathcal{Y}_j(t), \mathcal{D}_n)$ and we set $\text{EKL}(u \mid t) = 0$: no information acquired. This is a matter of choice, as (4.6) cannot be evaluated for $t < T_j^* < u$. On the other hand, when the event has not occurred up to u, the most information can be gained by setting u further away from the last measurement at t. The second term in (4.5) is a penalty term. Not only did $\text{EKL}(u \mid t)$ already penalize for scheduling measurement $u > T_j^*$, the risk term also bears in mind that the patients have to wait until u, thereby increasing the risk of an event. From a medical point of view, physicians would rather not wait until it is too late. The expectation of the indicator function yields the conditional probability of surviving up to u, given that the subject has already survived past t:

$$\pi_j(u \mid t) = p[T_j^* > u \mid T_j^* > t, \mathcal{Y}_j(t), \mathcal{D}_n]. \tag{4.7}$$

Predictions of conditional survival probabilities have been subject of research in Rizopoulos (2011). Estimates are based upon the posterior expectation of (4.7) and a detailed overview is given in the Appendix A.1.

As a matter of choice, the user can specify the non-negative weights λ_1 and λ_2, favoring either more information gain or heavier penalizing the risk function. Different choices of these weights will yield different optimal u. Clyde and Chaloner (1996) have shown that maximization of EKL(u_t) for any pair $(\lambda_1, \lambda_2), \exists \kappa \in [0, 1]$ such that the above is equivalent to maximizing EKL($u \mid t$) subject to $\pi_j(u \mid t) \geq \kappa$. Now, we need to specify a threshold k. Under this specification, κ could be set equal to the maximum risk of an event that the physician is willing to take. Now, we can search for the optimal value of u, where we maximize EKL($u \mid t$) in the interval $(t, t^{\text{up}}]$, where t^{up} is the 'limit time' of the survival process for threshold κ, i.e. $t^{\text{up}} = \min\{u : \pi_j(u \mid t) = \kappa\}$. If, on medical grounds, the physician necessarily wants to see patient j before time $t^{\max}$, then t^{up} needs to be censored by $t^{\max}$.

4.3 Estimation of the Kullback-Leibner divergence

The estimation of EKL($u \mid t$) makes use of the assumptions (3.2) and (3.16), in combination with the likelihood specification (3.20). The predictive distribution for the event-time T_j^* is

$$p(T_j^* \mid T_j^* > t, \{\mathcal{Y}_j(t), y_j(u)\}, \mathcal{D}_n) = \int_{\boldsymbol{\theta}} p(T_j^* \mid T_j^* > t, \{\mathcal{Y}_j(t), y_j(u)\}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathcal{D}_n) d\boldsymbol{\theta} \tag{4.8}$$

where the part under the integral is equal to

$$\begin{aligned} &p(T_j^* \mid T_j^* > t, \{\mathcal{Y}_j(t), y_j(u)\}, \boldsymbol{\theta}) = \\ &\int_{\boldsymbol{u}_j} p(T_j^* \mid T_j^* > t, \boldsymbol{u}_j, \boldsymbol{\theta}) p(\boldsymbol{u}_j \mid T_j^* > t, \{\mathcal{Y}_j(t), y_j(u)\}, \boldsymbol{\theta}) d\boldsymbol{u}_j \end{aligned} \tag{4.9}$$

and for the new longitudinal measurement at time u.

$$p(y_j(u) \mid T_j^* > t, \mathcal{Y}_j(t), \mathcal{D}_n) = \int_{\boldsymbol{\theta}} p(y_j(u) \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathcal{D}_n) d\boldsymbol{\theta} \tag{4.10}$$

which is simply the prediction of the biomarker at time u. When we now look at the same distribution but for $T_j^* > u$, we merely plug in the above for u. Concisely, the predictive distribution of an event later than u is given by

$$\begin{aligned} &p(T_j^* \mid T_j^* > u, \{\mathcal{Y}_j(t), y_j(u)\}, \mathcal{D}_n) \\ &= \int_{\boldsymbol{\theta}} \int_{\boldsymbol{u}_j} p(T_j^* \mid T_j^* > u, \boldsymbol{u}_j, \boldsymbol{\theta}) p(\boldsymbol{u}_j \mid T_j^* > u, \{\mathcal{Y}_j(t), y_j(u)\}, \boldsymbol{\theta}) d\boldsymbol{u}_j d\boldsymbol{\theta} \end{aligned} \tag{4.11}$$

and the risk-process under the integral above is given by

$$p(T_j^* \mid T_j^* > u, \boldsymbol{u}_j, \boldsymbol{\theta}) = \frac{\lambda_j(T_j^* \mid \mathcal{M}_j(t), \boldsymbol{u}_j, \boldsymbol{\theta}) S_j(T_j^* \mid \mathcal{M}_j(t), \boldsymbol{u}_j, \boldsymbol{\theta})}{S_j(u \mid \mathcal{M}_j(t), \boldsymbol{u}_j, \boldsymbol{\theta})}. \tag{4.12}$$

Based on the equations above, we obtain the following Monte Carlo simulation scheme to estimate EKL($u \mid t$) in (4.6):

1. Simulate $\tilde{\boldsymbol{\theta}}, \dot{\boldsymbol{\theta}}$ and $\{\boldsymbol{\theta}^{(l)}, l = 1, \ldots, L\} \sim p(\boldsymbol{\theta}|\mathcal{D}_n)$
2. Draw $\tilde{\boldsymbol{u}_j} \sim p(\boldsymbol{u}_j \mid T_j^* > t, \mathcal{Y}_j(t), \tilde{\boldsymbol{\theta}})$
3. Draw $\tilde{y}_j(u) \sim p(y_j(u) \mid \tilde{\boldsymbol{u}_j}, \tilde{\boldsymbol{\theta}})$
4. Simulate $\dot{\boldsymbol{u}_j}$ from $p(\boldsymbol{u}_j \mid T_j^* > t, \{\mathcal{Y}_j(t), \tilde{y}_j(u)\}, \dot{\boldsymbol{\theta}})$ and $\{\boldsymbol{u}_{j+}^{(l)}, l = 1, \ldots, L\}$ from $p(\boldsymbol{u}_j \mid T_j^* > u, \{\mathcal{Y}_j(t), \tilde{y}_j(u)\}, \dot{\boldsymbol{\theta}^{(l)}})$ and $\{\boldsymbol{u}_{j-}^{(l)}, l = 1, \ldots, L\}$ from $p(\boldsymbol{u}_j \mid T_j^* > u, \mathcal{Y}_j(t), \dot{\boldsymbol{\theta}^{(l)}})$.
5. Simulate $\dot{T}_j^*$ from $p(T_j^* \mid T_j^* > t, \dot{\boldsymbol{u}_j}, \dot{\boldsymbol{\theta}})$
6. When $\dot{T}_j^* > u$ we obtain estimates from $\mathrm{EKL}^{(q)}(u \mid t) = \log\left\{\frac{1}{L}\sum_{l=1}^{L} \mathcal{A}_n^{(l)}/\mathcal{A}_d^{(l)} \Big/ \mathcal{B}_n^{(l)}/\mathcal{B}_d^{(l)}\right\}$

   $\mathcal{A}_n^{(l)} = \lambda_j(\dot{T}_j^* \mid \mathcal{M}_j(\dot{T}_j^*), \boldsymbol{u}_{j+}^{(l)}, \boldsymbol{\theta}^{(l)}) S(\dot{T}_j^* \mid \mathcal{M}_j, \boldsymbol{u}_{j+}^{(l)}, \boldsymbol{\theta}^{(l)})$,
   $\mathcal{A}_d^{(l)} = S(u \mid \mathcal{M}_j, \boldsymbol{u}_{j+}^{(l)}, \boldsymbol{\theta}^{(l)})$.
   $\mathcal{B}_n^{(l)} = \lambda_j(\dot{T}_j^* \mid \mathcal{M}_j(\dot{T}_j^*), \boldsymbol{u}_{j-}^{(l)}, \boldsymbol{\theta}^{(l)}) S(\dot{T}_j^* \mid \mathcal{M}_j, \boldsymbol{u}_{j-}^{(l)}, \boldsymbol{\theta}^{(l)})$,
   $\mathcal{B}_d^{(l)} = S(u \mid \mathcal{M}_j, \boldsymbol{u}_{j-}^{(l)}, \boldsymbol{\theta}^{(l)})$.

   When $\dot{T}_j^* < u$, $\mathrm{EKL}^{(q)}(u \mid t) = 0$.
7. $\widehat{\mathrm{EKL}}(u \mid t) = \frac{1}{Q}\sum_{q=1}^{Q} \mathrm{EKL}^{(q)}(u \mid t)$

For the model parameters in $\boldsymbol{\theta}$ and $\boldsymbol{u}_i$ in Step 1, Step 2 and Step 4, we will sample from the posterior distribution given in (3.20). Since these distributions are non-standard, we propose a Metropolis-Hastings algorithm based on a multivariate Student's t-distribution, with mean $\hat{\boldsymbol{u}} = \arg\max_{\boldsymbol{u}}\{\log p(\boldsymbol{u} \mid T_j^* > t, \mathcal{Y}_j(t), \hat{\boldsymbol{\theta}}\}$. The variance-covariance matrix V is approximated by the Hessian $\{\partial^2 \log p(\boldsymbol{u} \mid T_j^* > t, \mathcal{Y}_j(t), \hat{\boldsymbol{\theta}})/\partial \boldsymbol{u}^\intercal \partial \boldsymbol{u}|_{\boldsymbol{u}=\hat{\boldsymbol{u}}}\}$. The posterior mean of the parameters $\boldsymbol{\theta}$ is denoted by $\hat{\boldsymbol{\theta}}$. In the third step, we merely make a prediction via the mixed model at time u. To draw an event time $\dot{T}_j^*$ after t, we use the cumulative density distribution $S_j(\dot{T}_j^*, \dot{\boldsymbol{u}_i}, \dot{\boldsymbol{\theta}})/S_j(t, \dot{\boldsymbol{u}_i}, \dot{\boldsymbol{\theta}})$ for inversion sampling. After having drawn $v \sim U(0, 1)$, we compute $\dot{T}_j^*$ for which the above CDF is equal to v. An estimate of the survival probabilities $\pi(u \mid t)$ helps us to find the optimal u: $u^{\mathrm{opt}} = \arg\max_u \widehat{\mathrm{EKL}}(u \mid t)$ subject to $\hat{\pi}(u \mid t) \geq \kappa$. In practice, we will assess a finite grid $\{u_1, \ldots, u_{\max}\}$ in the interval $(t, t^{\max}]$. We will propose different new scheduling times and choose the appointment time that yields highest information gain.

5 Extensions

Our extension of the work of Rizopoulos et al. (2016) is two-fold. First, we explore a Bayesian model averaged counterpart of $\mathrm{EKL}(u \mid t)$. Essentially, we adopt the previous

framework with ideas from Rizopoulos et al. (2014). Second, we extend the setting to allow for scheduling multiple screenings at once.

5.1 Bayesian model averaging

A possible drawback of the model selection routine based on the cvDCL (4.2) is that we fully focus our attention to one model: the highest-ranking in terms of cvDCL. The expected information ratio is hence also based solely on one model. As discussed by Hoeting et al. (1999), this approach does not take model uncertainty into account. To relax this restriction, we could apply the ideas of Bayesian model averaging (BMA) in a joint model framework. Following Hoeting et al. (1999), for a given quantity of interest Δ, the posterior distribution given the data equals

$$p(\Delta \mid \mathcal{D}) = \sum_{k=1}^{K} p(\Delta \mid \mathcal{D}, M_k) p(M_k \mid \mathcal{D}). \tag{5.1}$$

In our application, the quantity of interest is $\text{EKL}(u \mid t)$. We combine the ideas of Rizopoulos et al. (2014) and Rizopoulos et al. (2016) to come up with a relaxation of (4.6). Let subject j be a new subject in our dataset. The model average conditional survival probabilities of surviving up to time $u > t$ equal:

$$p(T_j^* > u \mid \mathcal{D}_j(t), \mathcal{D}_n) = \sum_{k=1}^{K} \underbrace{p(T_j^* > u \mid M_k, \mathcal{D}_j(t), \mathcal{D}_n)}_{\text{Model-specific survival probabilities}} \underbrace{p(M_k \mid \mathcal{D}_j(t), \mathcal{D}_n)}_{\text{Posterior weights}}. \tag{5.2}$$

The weights in the equation above still maintain a cross-validatory flavor, since they depend not only on the dataset $\mathcal{D}_n$, but also on the newly observed data of subject j, $\mathcal{D}_j(t)$. This results in subject and time-specific model selection, whereas the cvDCL indicates only time-specific model preference. In usual BMA parlance, the posterior weights can be obtained via:

$$p(M_k \mid \mathcal{D}_j(t), \mathcal{D}_n) = \frac{p(\mathcal{D}_j(t) \mid M_k) p(\mathcal{D}_n \mid M_k) p(M_k)}{\sum_{m=1}^{K} p(\mathcal{D}_j(t) \mid M_m) p(\mathcal{D}_n \mid M_m) p(M_m)}, \tag{5.3}$$

where

$$p(\mathcal{D}_j(t) \mid M_k) = \int_{\boldsymbol{\theta}_k} p(\mathcal{D}_j(t) \mid \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k \mid M_k) d\boldsymbol{\theta}_k \tag{5.4}$$

and the same holds for $p(\mathcal{D}_n \mid M_k)$. We note that

$$\begin{aligned} p(\mathcal{D}_n, \boldsymbol{u}_i, \boldsymbol{\theta}_k \mid M_k) &= \prod_{i=1}^{N} p(\boldsymbol{y}_i, T_i, \delta_i, \boldsymbol{u}_i, \boldsymbol{\theta}_k \mid M_k) \\ &= \prod_{i=1}^{N} p(y_i \mid \boldsymbol{u}_i, \boldsymbol{\theta}_k) p(T_i, \delta_i \mid \boldsymbol{u}_i, \boldsymbol{\theta}) p(\boldsymbol{u}_i \mid \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k \mid M_k) \end{aligned} \tag{5.5}$$

and by integrating out the random effects $\boldsymbol{u}_i$ and the model parameters $\boldsymbol{\theta}_k$, we obtain $p(\mathcal{D}_n \mid M_k)$. The likelihood $p(\mathcal{D}_n \mid \boldsymbol{\theta}_k)$ is the same as (3.21), i.e. just the likelihood evaluated under model k. For $p(\mathcal{D}_j(t) \mid \boldsymbol{\theta}_k)$ holds

$$p(\mathcal{D}_j(t) \mid \boldsymbol{\theta}_k) = p(\mathcal{Y}_j(t)|\boldsymbol{u}_j, \boldsymbol{\theta}_k) S_j(t \mid \boldsymbol{u}_j, \boldsymbol{\theta}_k) p(\boldsymbol{u}_j \mid \boldsymbol{\theta}_k) \tag{5.6}$$

and we explicitly state that $\boldsymbol{\theta}_k$ is fully estimated based on the dataset *without* subject j. We assume the a-priori model probabilities to be equal, i.e $p(M_k) = \frac{1}{K} \forall k$. Approximations of (5.5) and (5.6) and are obtained via Laplace approximations. For technical details, we refer to and Rizopoulos et al. (2014) and Rizopoulos et al. (2009).

To use the above in the context of personalized screening, we now adjust the Kullback-Leibner divergence accordingly. We will maximize

$$\begin{aligned} U(u \mid t) &= \sum_{m=1}^{K} \mathbb{E}\left\{ \log \frac{p(T_j^* \mid M_m, T_j^* > u, \{\mathcal{Y}_j(t), y_j(u)\}, \mathcal{D}_j(t), \mathcal{D}_n)}{p(T_j^* \mid M_m, T_j^* > u, \mathcal{Y}_j(t), \mathcal{D}_j(t), \mathcal{D}_n))} \right\} p(M_m \mid \mathcal{D}_j(t), \mathcal{D}_n) \\ \textbf{s.t. } \pi_j(u \mid t) &= \sum_{m=1}^{K} p(T_j^* > u \mid M_m, T_j^* > t, \mathcal{Y}_j(t), \mathcal{D}_j(t), \mathcal{D}_n) p(M_m \mid \mathcal{D}_j(t), \mathcal{D}_n) \geq \kappa. \end{aligned} \tag{5.7}$$

The restriction above is the averaged counterpart of (4.7). Similarly, the chance of an event occurring can not exceed κ, yet we now base the probability of occurrence on several models. To compute the above, we have to apply the sampling scheme of Section 4.3 for model $k = 1, \ldots, K$ times. Computationally, this might become a burden.

5.2 Scheduling multiple screenings

The ideas of Rizopoulos et al. (2016) can be extended to a setting with multiple screenings. The planning of multiple measurements is attractive from a time-management point of view, since the date(s) of the upcoming visits are known well in advance. We assume that the number of measurements to plan is decided upon by the physician and predetermined. Here, we will work out the simultaneous planning of two measurements: an extension to S measurements is straightforward and given in the Appendix A.2.

We again consider patient j at time point t, with history $\mathcal{D}_j(t)$. For this patient, we would like to plan two measurements at timepoints s_1, s_2. Without loss of generality, we assume $s_1 < s_2$. We will now look at the predictive posterior distribution with measurements $y_j(s_1), y_j(s_2)$. Since the physician decided to plan two measurements, we would ideally observe that the event-time T_j^* is later than s_2. We will adjust the utility function (4.5) to

$$U(s_1, s_2 \mid t) = \mathbb{E}\left\{ \lambda_1 \log \frac{p(T_j^* \mid T_j^* > s_2, \{\mathcal{Y}_j(t), y_j(s_1), y_j(s_2)\}, \mathcal{D}_n)}{p(T_j^* \mid T_j^* > s_2, \mathcal{Y}_j(t), \mathcal{D}_n)} - \lambda_2 I(T_j^* > s_2) \right\}. \tag{5.8}$$

In the setting of multiple measurements, we can trichotomize the real event time, T_j^*. The event occurs (i) before any measurement has taken place, in which case $T_j^* < s_1$

and we set $\text{EKL}(s_1, s_2 \mid t) = 0$, or (ii) after s_1, but before the second measurement s_2: $s_1 < T_j^* < s_2$, in which case we put $\text{EKL}(s_1, s_2 \mid t) = \text{EKL}(s_1 \mid t)$, or the event can take place (iii) after both measurements, i.e. $T_j^* > s_2$. In this way, the first term takes into account the true event time.

The last term again arises as a penalizing term, accounting for the risk of having to wait up till s_2. Via similar reasoning as in the previous section, we will optimize the information gain subject to a predetermined survival constraint based on the penalty term:

$$\pi_j(s_2 \mid t) = p(T_j^* > s_2 \mid T_j^* > t, \mathcal{Y}_j(t), \mathcal{D}_n) \tag{5.9}$$

where again, $\pi_j(s_2 \mid t) \geq \kappa$ is a necessary condition. To obtain a sampling scheme, we first look at the individual parts of the joint density $p(T_j^*, y_j(s_1), y_j(s_2) \mid T_j^* > t, \mathcal{Y}_j(t), \mathcal{D}_n)$. For T_j^*:

$$\begin{aligned} &p(T_j^* \mid T_j^* > t, \{\mathcal{Y}_j(t), y_j(s_1), y_j(s_2)\}, \mathcal{D}_n) = \\ &\int_{\boldsymbol{\theta}} p(T_j^* \mid T_j^* > t, \{\mathcal{Y}_j(t), y_j(s_1), y_j(s_2)\}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathcal{D}_n) d\boldsymbol{\theta} \end{aligned} \tag{5.10}$$

where the part under the integral is equal to

$$\begin{aligned} &p(T_j^* \mid T_j^* > t, \{\mathcal{Y}_j(t), y_j(s_1), y_j(s_2)\}, \boldsymbol{\theta}) = \\ &\int_{\boldsymbol{u}_j} p(T_j^* \mid T_j^* > t, \boldsymbol{u}_j, \boldsymbol{\theta}) p(\boldsymbol{u}_j \mid T_j^* > t, \{\mathcal{Y}_j(t), y_j(s_1), y_j(s_2)\}, \boldsymbol{\theta}) d\boldsymbol{u}_j. \end{aligned} \tag{5.11}$$

For the two new measurements holds:

$$p(y_j(s_1), y_j(s_2) \mid T_j^* > t, \mathcal{Y}_j(t), \mathcal{D}_n) = \int_{\boldsymbol{\theta}} p(y_j(s_1), y_j(s_2) \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathcal{D}_n) d\boldsymbol{\theta} \tag{5.12}$$

And the joint conditional distribution $p(y_j(s_1), y_j(s_2) \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta})$ can be written as

$$\begin{aligned} &p(y_j(s_1), y_j(s_2) \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}) = \\ &\int_{\boldsymbol{u}_j} p(y_j(s_1), y_j(s_2) \mid \boldsymbol{u}_j, \boldsymbol{\theta}) p(\boldsymbol{u}_j \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}) d\boldsymbol{u}_j = \\ &\int_{\boldsymbol{u}_j} \underbrace{p(y_j(s_1) \mid \boldsymbol{u}_j, \boldsymbol{\theta}) p(y_j(s_2) \mid \boldsymbol{u}_j, \boldsymbol{\theta})}_{\text{Conditional independence (3.2)}} p(\boldsymbol{u}_j \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}) d\boldsymbol{u}_j \end{aligned} \tag{5.13}$$

where the last step follows from the independence of longitudinal measurements, given the random effects. Now, the rest of the sampling scheme follows exactly the route from Section 4.3, with the only difference that we draw two longitudinal measurements at times s_1 and s_2.

1. Simulate $\tilde{\boldsymbol{\theta}}^1, \tilde{\boldsymbol{\theta}}^2, \dot{\boldsymbol{\theta}}$ and $\{\boldsymbol{\theta}^{(l)}, l = 1, \ldots, L\} \sim p(\boldsymbol{\theta}|\mathcal{D}_n)$
2. Draw $\tilde{\boldsymbol{u}_j}^1 \sim p(\boldsymbol{u}_j \mid T_j^* > t, \mathcal{Y}_j(t), \tilde{\boldsymbol{\theta}^1})$ and $\tilde{\boldsymbol{u}_j}^2 \sim p(\boldsymbol{u}_j \mid T_j^* > t, \mathcal{Y}_j(t), \tilde{\boldsymbol{\theta}}^2)$
3. Draw $\tilde{y}_j(s_1) \sim p(y_j(s_1) \mid \tilde{\boldsymbol{u}_j}^1, \tilde{\boldsymbol{\theta}}^1)$ and $\tilde{y}_j(s_2) \sim p(y_j(s_2) \mid \tilde{\boldsymbol{u}_j}^2, \tilde{\boldsymbol{\theta}}^2)$
4. Simulate $\dot{\boldsymbol{u}_j}$ from $p(\boldsymbol{u}_j \mid T_j^* > t, \{\mathcal{Y}_j(t), \tilde{y}_j(s_1), \tilde{y}_j(s_2)\}, \dot{\boldsymbol{\theta}})$ and $\{\boldsymbol{u}_{j+}^{(l)}, l = 1, \ldots, L\}$ from $p(\boldsymbol{u}_j \mid T_j^* > u, \{\mathcal{Y}_j(t), \tilde{y}_j(s_1), \tilde{y}_j(s_2\}, \dot{\boldsymbol{\theta}^{(l)}})$ and $\{\boldsymbol{u}_{j-}^{(l)}, l = 1, \ldots, L\}$ from $p(\boldsymbol{u}_j \mid T_j^* > u, \mathcal{Y}_j(t), \dot{\boldsymbol{\theta}^{(l)}})$
5. Simulate $\dot{T}_j^*$ from $p(T_j^* \mid T_j^* > t, \dot{\boldsymbol{u}_i}, \dot{\boldsymbol{\theta}})$
6. When $\dot{T}_j^* > s_2$ we obtain estimates from $\text{EKL}^{(q)}(s_1, s_2 \mid t) = \log\left\{\frac{1}{L}\sum_{l=1}^{L} \frac{\mathcal{A}_n^{(l)}}{\mathcal{A}_d^{(l)}} / \frac{\mathcal{B}_n^{(l)}}{\mathcal{B}_d^{(l)}}\right\}$

   $\mathcal{A}_n^{(l)} = \lambda_j(\dot{T}_j^* \mid \mathcal{M}_j(\dot{T}_j^*), \boldsymbol{u}_{j+}^{(l)}, \boldsymbol{\theta}^{(l)}) S(\dot{T}_j^* \mid \mathcal{M}_j, \boldsymbol{u}_{j+}^{(l)}, \boldsymbol{\theta}^{(l)})$,
   $\mathcal{A}_d^{(l)} = S(u \mid \mathcal{M}_j, \boldsymbol{u}_{j+}^{(l)}, \boldsymbol{\theta}^{(l)})$.
   $\mathcal{B}_n^{(l)} = \lambda_j(\dot{T}_j^* \mid \mathcal{M}_j(\dot{T}_j^*), \boldsymbol{u}_{j-}^{(l)}, \boldsymbol{\theta}^{(l)}) S(\dot{T}_j^* \mid \mathcal{M}_j, \boldsymbol{u}_{j-}^{(l)}, \boldsymbol{\theta}^{(l)})$,
   $\mathcal{B}_d^{(l)} = S(u \mid \mathcal{M}_j, \boldsymbol{u}_{j-}^{(l)}, \boldsymbol{\theta}^{(l)})$.

   When $\dot{T}_j^* < s_1$, $\text{EKL}^{(q)}(s_1, s_2 \mid t) = 0$.
   When $s_1 < \dot{T}_j^* < s_2$, $\text{EKL}^{(q)}(s_1, s_2 \mid t) = \text{EKL}^{(q)}(s_1 \mid t)$, see Section 4.3.
7. Repeat the above $q = 1, \ldots, Q$ times: $\widehat{\text{EKL}}(s_1, s_2 \mid t) = \frac{1}{Q}\sum_{q=1}^{Q} \text{EKL}^{(q)}(s_1, s_2 \mid t)$

6 Results

We now go back to the coronary disease dataset from the bioSHiFT studies in Section 1. Out of the 263 patients in bioSHiFT, 189 were males and 74 females. 70 (27 %) reached their primary endpoint: 53 (28%) men and 17(23%) women. The median age was 67 years and the median follow-up time was 2.17 years. Up to this point, 1929 blood samples were taken. We are interested in possible differences between men/women and or younger/older patients. Afterwards, we will plan new measurements for new patients.

We will start by introducing the set of joint models tested in our analysis. To model the longitudinal outcome, we included age, gender, and their interaction effect as explanatory variables. $\texttt{Age}_i$ is the age of patient i at baseline, centered around the median age of 67 years. $\texttt{Female}_i$ is a dummy variable for females measured at baseline, and $\texttt{FemAge}_i$ denotes the age/gender interaction. We made use of 2 flexible B-splines, denoted by $B_n(t, \{1, 2\})$.

$$\begin{aligned} y_i(t) =& (\beta_0 + u_{i0}) + (\beta_1 + u_{i1})B_n(t, 1) + (\beta_2 + u_{i2})B_n(t, 2) \\ &+ \beta_3 \texttt{Age}_i + \beta_4 \texttt{Female}_i + \beta_5 \texttt{FemAge}_\texttt{i} + \varepsilon_i(t) \end{aligned}$$

Following (3.13), we consider different postulations for the baseline hazards of the survival model:

$$M_1 : \lambda_i(t) = \lambda_0(t) \exp\{\gamma_1 \texttt{Age}_i + \gamma_2 \texttt{Female}_i + \gamma_3 \texttt{FemAge}_i + \alpha_1 m_i(t)\}$$

$$M_2 : \lambda_i(t) = \lambda_0(t) \exp\{\gamma_1 \texttt{Age}_i + \gamma_2 \texttt{Female}_i + \gamma_3 \texttt{FemAge}_i + \alpha_2 m_i'(t)\}$$

$$M_3 : \lambda_i(t) = \lambda_0(t) \exp\{\gamma_1 \texttt{Age}_i + \gamma_2 \texttt{Female}_i + \gamma_3 \texttt{FemAge}_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\}$$

$$M_4 : \lambda_i(t) = \lambda_0(t) \exp\{\gamma_1 \texttt{Age}_i + \gamma_2 \texttt{Female}_i + \gamma_3 \texttt{FemAge}_i + \alpha_3 \int_0^t m_i(s)ds\}$$

$$M_5 : \lambda_i(t) = \lambda_0(t) \exp\{\gamma_1 \texttt{Age}_i + \gamma_2 \texttt{Female}_i + \gamma_3 \texttt{FemAge}_i + \alpha_1 m_i(t) + \alpha_3 \int_0^t m_i(s)ds$$

and we will consider M_1 to be the standard model.

As parametrization, we specify diffuse priors for all regression coefficients, i.e. we take normal priors for the fixed effects $\boldsymbol{\beta}$, for the association parameter $\boldsymbol{\alpha}$ and for the parameters in the survival process $\boldsymbol{\gamma}$. All have mean 0 and variance 100. Rizopoulos et al. (2009) have shown that, as the number of longitudinal measurements increases, the mixed model contribution predominates the likelihood. For the variance σ^2 in the longitudinal outcome, we specify an inverse-Gamma 2 distribution with shape 6.1 from (3.1) and rate 0.78, for the variance-covariance matrix of the random effects $\boldsymbol{D}$ we specify an inverse-Wishart distribution with parameters $\hat{\boldsymbol{D}}$: the estimated variance-covariance matrix of only the longitudinal process (3.1) and 3 degrees of freedom.

6.1 Comparison of cvDCL and BMA

The cross-validated dynamical likelihoods (4.2) have been computed at 6 equidistant timepoints, and are displayed alongside the standard Bayesian model diagnostics DIC and LPML in Table 1. We choose to continue our analysis with the three best performing models. In the beginning, M_1 and M_5 rival as best performing models, with M_3 closing in in further measurements. All three perform comparable based on the cvDCL. Models M_2 and M_4 both underperform compared to the other three models. Furthermore, M_3 performs best according to DIC and LMPL criteria. Concludingly, we opt to work with M_1, M_3 and M_5.

Table 1: Comparison of $\widehat{\text{cvDCL}}(t)$ at 6 equidistant timepoints, DIC and LPML for fitted joint models

	M_1	M_2	M_3	M_4	M_5
t = 0.0	-182.23	-221.67	-191.26	-363.65	-137.63
t = 0.5	-143.92	-170.10	-148.55	-276.11	-111.18
t = 1.0	-100.68	-122.28	-104.62	-172.78	-83.65
t = 1.5	-66.90	-83.03	-69.02	-103.13	-58.59
t = 2.0	-29.45	-37.28	-30.15	-40.14	-27.67
t = 2.5	-6.42	-7.82	-6.22	-8.59	-6.11
DIC	5000.95	5044.49	5002.106	5028.392	5004.374
LPML	-2666.02	-2678.45	-2644.88	-2668.53	-2657.31

M_5 has the lowest $\widehat{\text{cvDCL}}(t)$ at all timepoints.

Parameter estimates of M_1 are depicted in Table 2. The posterior means of the parameters as well as 95% HPD-regions are shown. The association between the biomarker level α_1 and the log hazard function is 0.73. This implies that a unit change in biomarker level will, on average, result in a 2-fold (exp(0.73)) increase in hazard rate, thereby reducing the chances of survival. This is in line with earlier observations: high NT-proBNP levels for the patients who suffered an event. We observe that gender has no influence on the biomarker levels, as zero is contained in its 95% HPD-region. Every extra year aged from 67 increases the NT-proBNP by 0.05 on average. For the influence on the hazard we obtain: $\partial \frac{\lambda_i(t)}{\partial \texttt{Age}_i} = \gamma_{\texttt{Age}_i} + \alpha \times \beta_{\text{Age}_i}$. We assess the latter by the joint posterior distribution, giving a mean estimate of 0.067 with a 95% HPD region of $(0.038, 0.011)$.

Table 2: Posterior means and 95% HPD-regions of joint model M_1

	Survival				Longitudinal		
Variable	Mean	2.5%	97.5%		Mean	2.5%	97.5%
`Female`	-0.145	-0.40	0.75	`Intercept`	4.71	4.56	4.86
`Age`	0.03	-0.01	0.07	$B_n(t,1)$	0.05	-0.12	0.21
`FemAge`	-0.03	-0.07	-0.01	$B_n(t,2)$	0.25	0.11	0.39
α_1	0.73	0.53	0.93	`Female`	-0.07	-0.39	0.26
				`Age`	0.05	0.03	0.08
				`FemAge`	0.02	-0.04	0.04
				σ	0.36	0.34	0.37

As an illustration, we will further investigate the model outcomes for patient 23. We have chosen this patient, as the biomarker trajectory of patient 23 shows an interesting pattern. Patient 23 is a 48 year-old male that entered the screening process due to angina pectoris (chest pain). At the first screenings, the patient showed decreasing

levels of NT-proBNP. Later, NT-proBNP levels rise, indicating the deterioration of the patient's health. Based upon M_1, we show updated conditional survival probabilities in Figure 4. As the biomarker level decreases during the first 6 measurements, the curve of the mean conditional survival probabilities becomes more horizontal. From measurement 7 until 10 the level of NT-proBNP in the blood increases, resulting in a steep decrease of survival probability. Unfortunately for this patient, his chances of remaining event-free until 2.5 years after the start of the follow-up dropped from 90% at 1.1 years, to 50% at the last measurement within a year time.

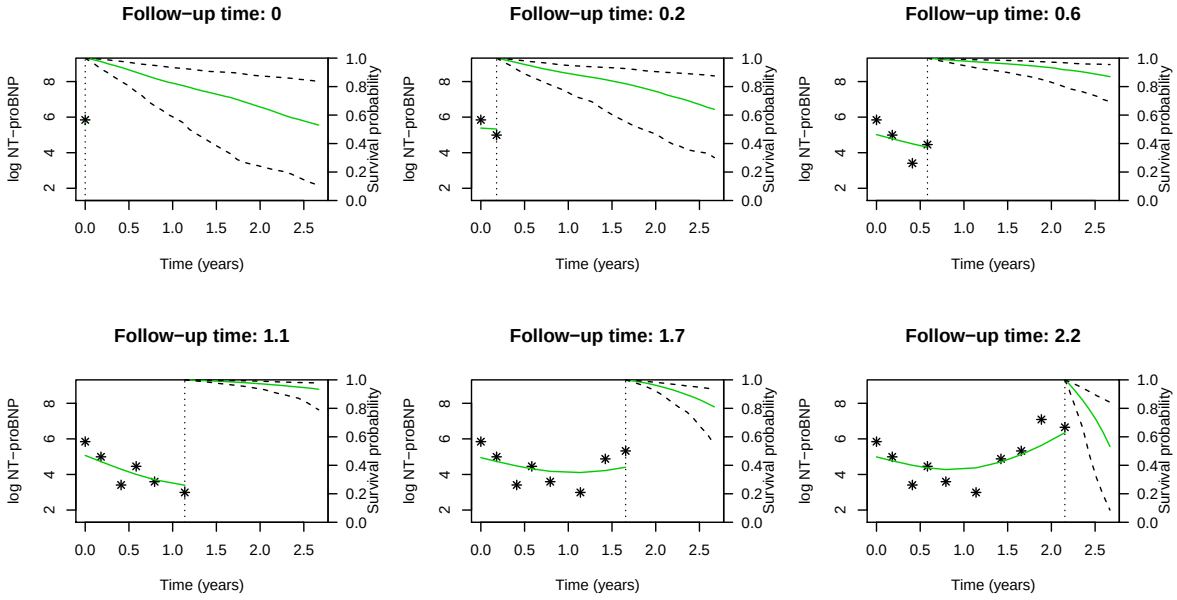

Figure 4: Dynamic conditional survival probabilities of patient 23. The longitudinal trajectory is displayed on the left. On the right, the mean survival is plotted in green. The dashed lines are 95% empirical confidence regions.

We will now continue to plan next measurements for patient 23. We compare the scheduling of the upcoming visits based on the model with the highest cvDCL and the Bayesian averaged counterpart. The cvDCL's and the posterior model probabilities can be found in Table 3. For every visit, M_5 is the best model according to the cvDCL. The differences between the cvDCL(t)'s are rather small, certainly if we compare this to the non-selected models M_2 and M_4 in Table 1. The posterior model probabilities tell another story. Here, we observe that M_3 most often has the highest posterior probability, although we observe shocks at 6 and 9. This is caused by the unexpected measurements at these times: NT-proBNP levels suddenly rose steeply. This shift indicates that for

this patient, the trajectory can perhaps not be modeled by one model alone. As a comparison, we have shown the results for a steady longitudinal trajectory of patient 19 in Appendix B.1. There, we observe no sudden shocks in model probabilities, indicating that one sole model in that case is sufficient to plan new measurements. The difference in choice of the two methods is partly based on the inclusion of longitudinal history of patient 23 in the BMA-procedure. Furthermore, the cvDCL accounts mainly for fit of the survival path, whereas BMA also assesses the modeling accuracy of the longitudinal trajectory.

Table 3: For every visit of patient 23, $\widehat{\text{cvDCL}}(t)$ on the left-hand side of the panel. The right-hand side depicts the posterior model probabilities.

	$\widehat{\text{cvDCL}}(t)$			Probability		
	M_1	M_3	M_5	M_1	M_3	M_5
t = 0.00	-182.23	-191.26	-137.63	0.13	0.48	0.39
t = 0.18	-174.70	-182.30	-130.93	0.12	0.55	0.33
t = 0.41	-148.11	-153.36	-113.44	0.32	0.59	0.09
t = 0.58	-135.22	-139.37	-105.52	0.00	0.93	0.07
t = 0.79	-115.96	-119.00	-92.99	0.00	0.97	0.03
t = 1.14	-91.81	-94.59	-77.13	0.00	0.97	0.03
t = 1.42	-69.00	-71.25	-59.64	0.02	0.06	0.92
t = 1.66	-51.61	-52.81	-45.66	0.16	0.65	0.19
t = 1.89	-37.37	-38.44	-34.35	0.90	0.03	0.07
t = 2.15	-15.11	-15.40	-14.96	0.58	0.16	0.26

For every visit, M_5 has the lowest $\widehat{\text{cvDCL}}(t)$.

We now compare the scheduling results based upon the best-performing model in terms of cvDCL, model M_5, to a scheduling based upon the posterior model probabilities. We first simulate survival chances $\pi_{23}(u \mid t)$ of patient 23. The doctor wants to see the patient yearly and hence the largest time frame allowed for the next measurement, called t^{max}, is equal to 1 year. As Rizopoulos et al. (2016), we do not want chances of survival dropping below 80%, and hence we set $\kappa = 0.8$. Within the assessed time frame, we compute $\text{EKL}(u \mid t)$ for 5 different u. The timepoint u with the highest $\text{EKL}(u \mid t)$ is then selected as the new measurement time.

The outcomes are depicted in Table 4. For the first six measurements, both models roughly predict the same survival probabilities and the interval for the next possible meeting is equal to the largest interval the physician is willing to wait. We observe similar selected times for the upcoming measurements, although the times selected by the model-averaged $\text{EKL}(u \mid t)$ seems to be more conservative. In particular for the last 3 visits, where the deterioration of patient 23 is also indicated by the shortening of the scheduling interval.

Table 4: On the left, we find the values of EKL($u \mid t$) and $\pi(u \mid t)$ for patient 23, based on M_5. On the right hand side, we find the model-averaged EKL($u \mid t$) and $\pi(u \mid t)$ of Section 5.1. The upper limit of the planning interval is $t^{\max}$. The optimal next screening time u_{opt} is denoted by *.

			cvDCL(t)				Bayesian Model Average		
t	$t^{\max}$	times	EKL($u \mid t$)	$\pi_j(u \mid t)$	*	$t^{\max}$	times	EKL($u \mid t$)	$\pi_j(u \mid t)$	*
0	1	0.20	-0.07	0.96		0.96	0.19	0.23	0.97	
		0.40	1.01	0.92	*		0.38	0.07	0.93	
		0.60	0.43	0.88			0.57	0.29	0.89	*
		0.80	0.65	0.83			0.77	0.24	0.85	
		1.00	0.13	0.79			0.96	0.21	0.81	
0.18	1	0.38	0.34	0.97		1	0.38	0.08	0.97	
		0.58	0.18	0.94			0.58	-0.10	0.92	
		0.78	0.30	0.90			0.78	0.29	0.88	*
		0.98	0.35	0.88	*		0.98	0.14	0.84	
		1.18	0.21	0.85			1.18	0.24	0.80	
0.41	1	0.61	0.18	0.98		1	0.61	0.38	0.97	
		0.81	0.35	0.97			0.81	0.37	0.92	
		1.01	0.33	0.96			1.01	0.40	0.88	*
		1.21	0.17	0.95			1.21	0.31	0.84	
		1.41	0.35	0.94	*		1.41	0.18	0.80	
0.58	1	0.78	0.28	0.98		1	0.78	0.41	0.97	
		0.98	0.29	0.97			0.98	0.31	0.92	
		1.18	0.48	0.96	*		1.18	0.44	0.88	*
		1.38	0.28	0.95			1.38	0.20	0.84	
		1.58	-0.08	0.94			1.58	0.25	0.80	
0.79	1	0.99	0.42	0.99		1	0.99	0.38	0.97	
		1.19	0.23	0.98			1.19	0.27	0.92	
		1.39	0.35	0.97			1.39	0.26	0.88	
		1.59	0.44	0.96	*		1.59	0.45	0.84	*
		1.79	0.42	0.95			1.79	0.44	0.80	
1.14	1	1.34	0.36	0.99		1	1.34	0.30	0.97	
		1.54	0.15	0.98			1.54	0.59	0.92	
		1.74	0.39	0.98			1.74	0.60	0.88	*
		1.94	0.03	0.96			1.94	0.27	0.84	
		2.14	0.55	0.95	*		2.14	0.36	0.80	
1.42	1	1.62	0.33	0.99		1	1.62	0.31	0.97	
		1.82	0.52	0.97	*		1.82	0.35	0.92	
		2.02	0.34	0.95			2.02	0.52	0.88	*
		2.22	0.28	0.93			2.22	0.19	0.84	
		2.42	0.43	0.90			2.42	0.51	0.80	
1.66	0.95	1.85	0.23	0.98		0.78	1.81	0.47	0.97	*
		2.04	0.36	0.95	*		1.97	0.42	0.92	
		2.23	0.22	0.91			2.12	0.43	0.88	
		2.42	0.24	0.87			2.28	0.33	0.84	
		2.61	0.32	0.81			2.44	0.30	0.80	
1.89	0.45	1.98	0.12	0.97		0.34	1.93	0.22	0.97	
		2.07	0.02	0.94			1.98	0.24	0.92	
		2.16	0.01	0.90			2.03	0.15	0.88	
		2.25	0.15	0.86	*		2.08	0.35	0.84	*
		2.34	0.02	0.81			2.13	0.18	0.80	
2.15	0.28	2.21	0.18	0.97	*	0.13	2.18	0.18	0.97	
		2.26	0.14	0.93			2.20	0.06	0.92	
		2.32	0.13	0.89			2.23	0.15	0.88	
		2.37	0.11	0.85			2.26	0.23	0.84	*
		2.43	0.01	0.81			2.28	0.22	0.80	

To compare the two methods, we need to revert to the theoretical framework. The cvDCL is based on posterior predictive power, thereby focusing mainly on the time-to-event outcome. Robustness is ensured by the cross-validatory aspect. A possible

drawback is the fact that the model choice is time - yet not patient-specific. For the Bayesian model averaging, we take the full model fit as a measure, and need not restrict ourselves to the selection of one sole model. As one of the steps in the sampling scheme of Section 4.3 is a realization from the mixed model, the use of the whole model fit is intuitive. A negative aspect is that the posterior model odds are very subjective to perturbation, i.e. a slight altering of the data for patient j causes a large shift in posterior model probabilities. Arguably, the method of choice is a matter of taste. For this particular patient 23, we see the benefits of using multiple models as there does not seem to be one model that accurately captures his trajectory. Computationally, the cvDCL-based selection is preferred to the Bayesian model averaging. For patient 23 with 10 visits, the computation time for the first method was approximately 25.3 minutes, where the BMA took 58.9 minutes on a computer with 16GB RAM and Intel Core i7 processor.

6.2 Illustration of two upcoming measurements

We now shift our focus to the simultaneous planning of 2 measurements, an illustration of the extension of Section 5.2. We assess a grid of 5 possible timepoints for the first measurement and 5 for the second measurement, thereby considering 25 combinations of treatment times. Again, we do not want the patient's estimated survival probabilities to drop below 80% before the last treatment, hence $\pi(s_2|t) > \kappa = 0.80$. The combination of highest estimated $\text{EKL}(s_1, s_2 \mid t)$ under the latter restriction is then scheduled. Computations are performed under M_5. The results are depicted in Table 5. The two measurements are scheduled at times $(0.38, 0.77)$. This means that, given the first biomarker measurement, we expect to gain the most insight in the trajectory at these given timepoints. In real-life, already 4 measurements had taken place at timepoint 0.79.

Table 5: $\text{EKL}(s_1, s_2 \mid t)$ of two measurements for patient 23 at his first visit.

t = 0	s_2	0.57	0.67	0.77	0.86	0.96
s_1						
0.10		0.06	0.09	0.08	-0.12	-0.13
0.19		0.13	0.13	0.15	-0.02	0.32
0.29		-0.04	0.41	0.11	-0.07	0.01
0.38		0.54	0.33	0.51	0.64	0.21
0.48		-0.01	0.34	**0.73**	0.37	0.28

The rows of the tabular correspond to the first measurement time s_1, the columns to the second measurement time s_2. The highest $\text{EKL}(s_1, s_2 \mid t)$ is depicted in bold.

To compare our results, we now look at a visit near the end of our follow-up for patient 23 in Table 6. At the seventh visit, NT-proBNP level has quickly started to rise, indicative

of health worsening. The EKL appears to show a certain pattern across the screening times. We observe that the left upper triangular part of the EKL-matrix roughly has higher values than the lower triangle: the model favors two screenings at short notice. Of course, investigating 25 different screening time combinations in less than half a year seems futile. The results, however, seem to draw similar conclusions as the planning of only one screening for this timepoint: the patient should be screened early.

Table 6: EKL$(s_1, s_2 \mid t)$ of two measurements for patient 23 at his seventh visit.

t=1.66	s_2	2.23	2.33	2.42	2.52	2.61
s_1						
1.75		0.07	0.05	0.29	-0.05	-0.45
1.85		0.17	0.31	**0.33**	0.33	0.13
2.94		0.10	0.27	0.30	-0.20	0.05
2.04		0.10	0.15	-0.17	-0.03	-0.19
2.13		-0.26	-0.19	-0.10	-0.25	-0.50

The rows of the tabular correspond to the first measurement time s_1, the columns to the second measurement time s_2. The highest EKL$(s_1, s_2 \mid t)$ is depicted in bold.

The computation time of planning of two measurements, for all visits for patient 23, with a 5×5 grid was 33.1 minutes. The remaining results can be found in Table 8 of Appendix B.2 . Not all visits display a similar structure as the seventh visit, where early screening was preferred. On the contrary, the exact optimal scheduling times are vague. This is partly perhaps due to over-planning: 25 possible measurements is a lot in a short time period of one year. In theory, we could also combine the Bayesian framework in a multiple screening setting. However, computing time will rise exponentially as the number of grid-points and models increases.

7 Discussion

We made an attempt to provide optimized personalized screening intervals. Optimality, however, is not uniquely defined. In this thesis, we considered a model to be optimal if it maximized the utility function (4.5), based on an information criterion. The choice of utility function therefore determines the best performing model - subject to the utility of choice. Different concepts of optimality, or utility functions to be optimized, lead to different optimal models. This raises a more philosophical question: how should optimality in screening procedures be defined?

The main purpose of screenings is to reduce morbidity and mortality, whilst bearing in mind the patient's burden and health care cost (Parmigiani, 1993). This requires a careful balance between both sides - all parties want to reassure early disease detection, but have different views of burden. Breast cancer screenings are not patient-friendly,

but patients do not want to be at great risk of breast cancer either. Moreover, there ought to be an incentive to perform screenings from a medical point of view, expressed in adjusted-quality of life years and treatment cost. One is most sure about a biomarker trajectory when measured as often as possible, but the cost of these measurements is simply too high - let alone the patient's cost to travel to the hospital every day. Furthermore, chances of detection need to be reasonably high: breast cancer screenings are only performed for 50+-year old women. Disease detection for women younger than 50 years does not outweigh the cost and unpleasantness of screenings (Lee and Zelen, 1998). A cost-effectiveness study weighs all advantages and disadvantages mentioned above. For an exemplary study in the field of screening for cervical cancer, see Rosmalen et al. (2012).

Ideally, we would optimize a criterion that takes all of the above into account - a cost-effective measure. Our criterion takes a reasonable shot and could be adjusted to take into account screening cost - that is simply penalizing for the number of screenings. The number of screenings itself should, however, also be subject to optimization. To perfection, we would optimize over the screening times and the number of screenings dynamically. Dynamically, since we would use all planned measurements over time and not jointly plan multiple measurements. When working with Markov-models, Parmigiani (1993) has shown that, via Bellman equations, this can lead to analytically solvable problems. Unfortunately, the transitions probabilities and states do not incorporate the unique feature of the joint modeling framework: jointly, personalized modeling longitudinal and survival outcomes. To derive a stochastic dynamical optimization routine in the field of joint models and personalized screening would be ideal, but is beyond the scope of this thesis.

8 Conclusion

In this thesis, we have taken a closer look at the biomarker NT-proBNP for heart failure patients. By jointly modeling both the trajectory of NT-proBNP levels in blood plasma and the effect of this trajectory on the patient's health status, we attempted to optimally plan a new measurement. We found that higher levels of NT-proBNP resulted in a risk-increase of cardiac events. The framework of personalized screening intervals of Rizopoulos et al. (2016) was the starting point to define optimality and thereupon investigate future screening times. We have focused on two limitations of their framework: (i) the model selection procedure that was based on one model solely and (ii) the planning of only one measurement ahead.

For the former limitation, we assessed a Bayesian model averaging framework to allow for combinations of models to specify the biomarker trajectory. Especially for patients with a non-standard evolution of biomarker levels, more flexibility arises to model the progression of NT-proBNP. In a real-life case for such a patient, we have seen the advantages of Bayesian model averaging in the joint modeling framework. A possible drawback for this method is the computational burden.

For the latter limitation, we extended the existing screening planning procedure for

one screening to a setting with possible S screenings. We opted to simultaneously plan these measurements, leading to S-step ahead screenings. We illustrated this specification with an exemplary patient, where we planned two upcoming visits. The joint planning of multiple screenings yields scheduling advantages for both patient and physician. A disadvantage is that this method is not always able to discriminate between measurement combinations, i.e. there does not always seem to be a clear preference regarding the exact optimal screening times.

In both cases, we worked with optimality as defined by Rizopoulos et al. (2016). As discussed in the previous section, it is in this direction that we feel lies the most room for future research. For example, by combining this thesis, Rizopoulos et al. (2016) and Parmigiani (1993, 1998, 2002). To us, this would result in adapting the existing framework for optimality in cost-effective ways and thereafter planning new screenings in a dynamical manner.

In the end, statisticians are no physicians - but physicians are no statisticians either. The field of personalized medicine requires a combination of both forms of expertise. As more and more patient information is registered and stored, statisticians can help physicians with data-analysis. This is not only helpful within hospitals, but nationwide or even internationally. We hope that this thesis is an example of an application of statistics/econometrics in the field of medicine. Of course, whether statistician or physician, the final goal is always the improvement of life quality for patients.

References

M. Clyde and K. Chaloner. The equivalence of constrained and weighted designs in multiple objective design problems. *Journal of the American Statistical Association*, 91:1236–1244, 1996.

T. Cover and J. Thomas. *Elements of Information Theory*. New York: Wiley, 1991.

P. Eilers and B. Marx. Flexible smoothing with b-splines and penalties. *Statistical Science*, 11:89–121, 1996.

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

T. R. Fleming and D. P. Harrington. *Counting Processes and Survival Analysis*. John Wiley and Sons, Inc., 1991.

Hartstichting. Hart-en vaatziekten in nederland. cijfers of kwaliteit van leven, ziekte en sterfte. 2014.

D. A. Harville. Bayesian inference for variance components using only error contrasts. *Biometrika*, 61:383–385, 1974.

J.A. Hoeting, D. Madigan, A.E. Raftery, and C. T. Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, 14:382–417, 1999.

N. Laird and J. Ware. Random-effects models for longitudinal data. *Biometrics*, 38: 963–974, 1982.

S. Lee and M. Zelen. Scheduling periodic examinations for the early detection of disease: Applications to breast cancer. *Journal of the American Statistical Association*, 88: 622–628, 1998.

P. Libby. Atherosclerosis: disease biology affecting the coronary vasulature. *American Journal of Cardiology*, 98:3:9, 2006.

C. Mueller. Biomarkers and acute coronary syndromes: an update. *European Heart Journal*, 35:552–556, 2014.

M.A. Newton and A.E. Raftery. Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B.*, 56:3–48, 1994.

G. Parmigiani. On optimal screening ages. *Journal of the American Statistical Association*, 88:622–628, 1993.

G. Parmigiani. Designing observation times for interval censored data. *Sankhya: The Indian Journal of Statistics*, 60:446–458, 1998.

G. Parmigiani. *Modeling in Medical Decision Making: A Bayesian Approach.* New York: Wiley, 2002.

D. Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67:819–829, 2011.

D. Rizopoulos. *Joint Models for Longitudinal and Time-to-Event Data.* Boca Raton: Chapman Hall CRC, 2012.

D. Rizopoulos, G. Verbeke, and E. Lesaffre. Fully exponential laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society, Series B*, 71:637–654, 2009.

D. Rizopoulos, A. Hatfield L, B. P. Carlin, and J.M. Takkenberg. Combining dynamic predictions from joint models for longitudinal and time-to-event data using bayesian model averaging. *Journal of the American Statistical Association*, 109:1385–1397, 2014.

D. Rizopoulos, J.M.G. Taylor, J. Van Rosmalen, E. W. Steyerberg, and J.J.M. Takkenberg. Personalized screening intervals for biomarkers using joint models for longitudinal and survival data. *Biostatistics*, 17:149–164, 2016.

J. Van Rosmalen, I.M.C.M. de Kok, and M. Van Ballegooijen. Cost-effectiveness of cervical cancer screening:cytology versus human papillomavirus dna testing. *BJOG*, 67:699–709, 2012.

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

G. Verbeke and G. Molenberghs. *Linear Mixed Models for Longitudinal Data.* Springer Series in Statistics. Springer-Verlag, 2000.

R.L. Wolpert and S.C. Schmidler. α-stable limit laws for harmonic mean estimators of marginal likelihoods. *Statistica Sinica*, 59:10–45, 2011.

A Appendix I

A.1 Predicted survival probabilities

Estimates of $\pi_j(u \mid t)$ are based upon the posterior distribution:

$$\begin{aligned}&\Pr[T_j^* > u \mid T_j^* > t, \mathcal{Y}_j(t), \mathcal{D}_n] \\ &\quad = \int_{\boldsymbol{\theta}} \Pr[T_j^* > u \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}] p(\boldsymbol{\theta} \mid \mathcal{D}_n) d\boldsymbol{\theta} \end{aligned} \tag{A.1}$$

where the first part of the integral can be worked out as

$$\begin{aligned}&\Pr[T_j^* > u \mid T_j^* > t, \mathcal{Y}_j(t), \mathcal{D}_n] \\ &\quad = \int_{\boldsymbol{u}_j} \Pr[T_j^* > u \mid T_j^* > t, \boldsymbol{u}_j, \boldsymbol{\theta}] p(\boldsymbol{u}_j \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}) d\boldsymbol{u}_i \\ &\quad = \int_{\boldsymbol{u}_j} \frac{S_j(u \mid \mathcal{M}_j(u), \boldsymbol{u}_j, \boldsymbol{\theta})}{S_j(t \mid \mathcal{M}_j(t), \boldsymbol{u}_j, \boldsymbol{\theta})} p(\boldsymbol{u}_j \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}) d\boldsymbol{u}_j. \end{aligned} \tag{A.2}$$

Now we can base our estimates on a Monte Carlo scheme, where we can draw from $p(\boldsymbol{\theta} \mid \mathcal{D}_n)$ as given in (3.20). The Monte Carlo scheme to estimate $\pi_j(u \mid t)$ is

1. Draw $\boldsymbol{\theta}^{(l)} \sim p(\boldsymbol{\theta} \mid \mathcal{D}_n)$.
2. Draw $\boldsymbol{u}_j^{(l)} \sim p(\boldsymbol{u}_j \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}^{(l)})$.
3. Compute $\hat{\pi}_j^{(l)}(u \mid t) = \frac{S_j(u \mid \mathcal{M}_j(u), \boldsymbol{u}_j^{(l)}, \boldsymbol{\theta}^{(l)})}{S_j(t \mid \mathcal{M}_j(t), \boldsymbol{u}_j^{(l)}, \boldsymbol{\theta}^{(l)})}$.
4. Repeat steps 1-3 L times for each subject j.

A.2 Scheduling S screenings

We extend the planning of 2 screenings to a setting of $\boldsymbol{S} = s_1, \ldots, s_S$ screening times. This is a straightforward modification of Section 5.2. Let, without loss of generality, $s_1 < s_2 < \ldots < s_S$ and let $\boldsymbol{S}_q, q \in 1, \ldots, S = \{s_1, \ldots s_q\}$ and $\boldsymbol{S}_{-q} = \{s_{q+1}, s_{q+2}, \ldots, s_S\}$. We aim to plain S screenings, so we want our patient to live at least up to time s_S. Our utility function becomes:

$$U(\boldsymbol{S} \mid t) = \mathbb{E}\left\{ \lambda_1 \log \frac{p(T_j^* \mid T_j^* > s_S, \{\mathcal{Y}_j(t), \boldsymbol{y}_j(\boldsymbol{S})\}, \mathcal{D}_n)}{p(T_j^* \mid T_j^* > s_s, \mathcal{Y}_j(t), \mathcal{D}_n)} - \lambda_2 I(T_j^* > s_s) \right\}, \tag{A.3}$$

where in (A.3) $\boldsymbol{y}_j(\boldsymbol{S}) = \{\cup_{s=1}^{S} y_j(s_s)\}$. Again, this is the information with all new measurements $\boldsymbol{S}$, compared to no measurement planned. Similar to the planning of two measurements, the patient could not survive up to s_S, i.e $T_j^* < s_q, q \in S$. Let, in this case, inspect the event time T_j^*. We adjust $EKL(\boldsymbol{S} \mid t)$, as when $T_j^* < s_1$ no screening has occurred and we set the information ratio to 0. When $s_q < T_j^* < s_{q+1} q \in S$, we

set $\text{EKL}(\boldsymbol{S} \mid t) = \text{EKL}(\boldsymbol{S}_q \mid t)$. In doing so, we account for the fact that although the patient did not live up to s_S, we still have information based on the measurements $s_1, \ldots, s_q$. Therefore, we do not have zero information gain, which is the case when the patient does not live until the planning of the first measurement.

For the sampling scheme, we will have to adjust the joint distribution of the screening outcomes:

$$p(\boldsymbol{y}_j(\boldsymbol{S}) \mid T_j^* > t, \mathcal{Y}_j(t), \mathcal{D}_n) = \int_{\boldsymbol{\theta}} p(\boldsymbol{y}_j(\boldsymbol{S}) \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathcal{D}_n) d\boldsymbol{\theta} \tag{A.4}$$

And the joint conditional distribution $p(\boldsymbol{y}_j(\boldsymbol{S}) \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta})$ can be written as

$$\begin{aligned}
&p(\boldsymbol{y}_j(\boldsymbol{S}) \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}) = \\
&\int_{\boldsymbol{u}_j} p(\boldsymbol{y}_j(\boldsymbol{S}) \mid \boldsymbol{u}_j, \boldsymbol{\theta}) p(\boldsymbol{u}_j \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}) d\boldsymbol{u}_j = \\
&\int_{\boldsymbol{u}_j} \underbrace{\left[\prod_{s=1}^{S} p(y_j(s_s) \mid \boldsymbol{u}_j, \boldsymbol{\theta}) \right]}_{\text{Conditional independence (3.2)}} p(\boldsymbol{u}_j \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}) d\boldsymbol{u}_j.
\end{aligned} \tag{A.5}$$

The above is an immediate result of (5.12). The sampling scheme now becomes:

1. Simulate $\tilde{\boldsymbol{\theta}}^s, s = 1, \ldots, S, \dot{\boldsymbol{\theta}}$ and $\{\boldsymbol{\theta}^{(l)}, l = 1, \ldots, L\} \sim p(\boldsymbol{\theta} | \mathcal{D}_n)$
2. Draw $\tilde{\boldsymbol{u}_j}^s \sim p(\boldsymbol{u}_j \mid T_j^* > t, \mathcal{Y}_j(t), \tilde{\boldsymbol{\theta}^s})$ for every measurement s.
3. Draw $\tilde{y}_j(s_s) \sim p(y_j(s_s) \mid \tilde{\boldsymbol{u}_j}^s, \tilde{\boldsymbol{\theta}}^s) \forall s \in S$
4. Simulate $\dot{\boldsymbol{u}_j}$ from $p(\boldsymbol{u}_j \mid T_j^* > t, \{\mathcal{Y}_j(t), \tilde{\boldsymbol{y}}_j(\boldsymbol{S}), \dot{\boldsymbol{\theta}})$ and $\{\boldsymbol{u}_{j+}^{(l)}, l = 1, \ldots, L\}$ from $p(\boldsymbol{u}_j \mid T_j^* > u, \{\mathcal{Y}_j(t), \tilde{\boldsymbol{y}}_j(\boldsymbol{S})\}, \dot{\boldsymbol{\theta}^{(l)}})$ and $\{\boldsymbol{u}_{j-}^{(l)}, l = 1, \ldots, L\}$ from $p(\boldsymbol{u}_j \mid T_j^* > u, \mathcal{Y}_j(t), \dot{\boldsymbol{\theta}^{(l)}})$
5. Simulate $\dot{T_j^*}$ from $p(T_j^* \mid T_j^* > t, \dot{\boldsymbol{u}_i}, \dot{\boldsymbol{\theta}})$
6. When $\dot{T_j^*} > s_S$ we obtain estimates from $\text{EKL}^{(m)}(\boldsymbol{S} \mid t) = \log \left\{ \frac{1}{L} \sum_{l=1}^{L} \frac{\mathcal{A}_n^{(l)}}{\mathcal{A}_d^{(l)}} / \frac{\mathcal{B}_n^{(l)}}{\mathcal{B}_d^{(l)}} \right\}$

   $\mathcal{A}_n^{(l)} = \lambda_j(\dot{T_j^*} \mid \mathcal{M}_j(\dot{T_j^*}), \boldsymbol{u}_{j+}^{(l)}, \boldsymbol{\theta}^{(l)}) S(\dot{T_j^*} \mid \mathcal{M}_j, \boldsymbol{u}_{j+}^{(l)}, \boldsymbol{\theta}^{(l)})$,
   $\mathcal{A}_d^{(l)} = S(u \mid \mathcal{M}_j, \boldsymbol{u}_{j+}^{(l)}, \boldsymbol{\theta}^{(l)})$.
   $\mathcal{B}_n^{(l)} = \lambda_j(\dot{T_j^*} \mid \mathcal{M}_j(\dot{T_j^*}), \boldsymbol{u}_{j-}^{(l)}, \boldsymbol{\theta}^{(l)}) S(\dot{T_j^*} \mid \mathcal{M}_j, \boldsymbol{u}_{j-}^{(l)}, \boldsymbol{\theta}^{(l)})$,
   $\mathcal{B}_d^{(l)} = S(u \mid \mathcal{M}_j, \boldsymbol{u}_{j-}^{(l)}, \boldsymbol{\theta}^{(l)})$.

   When $\dot{T_j^*} < s_1$, $\text{EKL}^{(m)}(\boldsymbol{S} \mid t) = 0$.
   When $s_q < \dot{T_j^*} < s_{q+1}$, $\text{EKL}^{(m)}(\boldsymbol{S} \mid t) = \text{EKL}^{(q)}(\boldsymbol{S}_q \mid t)$, see above.
7. Repeat the above $m = 1, \ldots, M$ times: $\widehat{\text{EKL}}(\boldsymbol{S} \mid t) = \frac{1}{M} \sum_{m=1}^{M} \text{EKL}^{(m)}(\boldsymbol{S} \mid t)$.

B Appendix II

B.1 Results of patient 19

The fitted longitudinal trajectory of the biomarker and the actual measurements of patient 19 are depicted in Figure 5. Similar to the results of patient 23, we have also displayed the conditional survival probabilities.

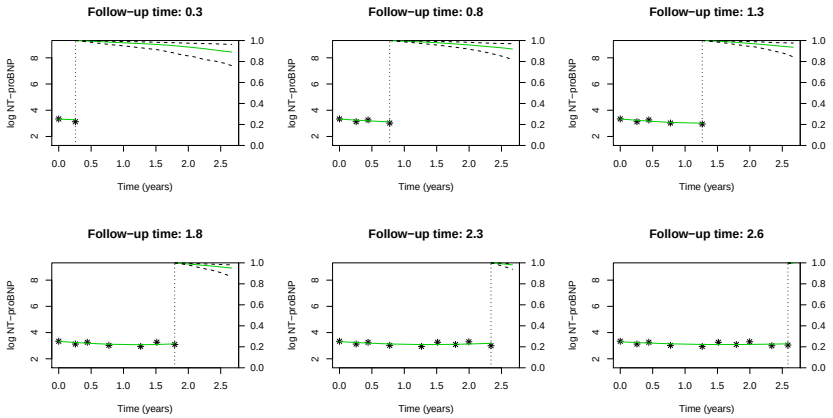

Figure 5: Dynamic conditional survival probabilities of patient 19. The longitudinal trajectory is displayed on the left. On the right, the mean survival is plotted in green. The dashed lines are 95% empirical confidence regions.

The posterior model probabilities are depicted in Table 7. At every visit, M_3 has the highest posterior probability. This is explained by the relatively stable trajectory of log NT-proBNP as depicted in Figure 5.

Table 7: Posterior model probabilities of patient 19, for every visit.

	M_1	M_3	M_5
t = 0.00	0.11	0.62	0.28
t = 0.18	0.01	0.60	0.39
t = 0.41	0.09	0.59	0.32
t = 0.77	0.08	0.59	0.33
t = 1.26	0.04	0.74	0.22
t = 1.51	0.09	0.70	0.21
t = 1.79	0.08	0.69	0.24
t = 2.00	0.08	0.75	0.17
t = 2.34	0.14	0.72	0.14
t = 2.59	0.00	0.82	0.18

B.2 EKL$(s_1, s_2 \mid t)$ for patient 23

Table 8: EKL$(s_1, s_2 \mid t)$ for patient 23, computed under M_5. All but visit 1 and 7 are displayed. At every visit, we considered 25 (5×5) screening time combinations, under the restriction that $\pi_{23}(s_1, s_2) > 0.8$. The rows of the tabular correspond to the first measurement time s_1, the columns to the second measurement time s_2. The upcoming scheduled screening is depicted in bold.

t=0.18	s_2	0.78	0.88	0.98	1.08	1.18
s_1						
0.28		0.14	0.49	**0.72**	0.15	0.32
0.38		0.49	0.37	0.23	0.62	0.13
0.48		0.21	0.34	0.44	0.21	0.12
0.58		0.40	0.14	-0.03	0.22	0.13
0.68		0.61	0.30	0.34	0.01	-0.03

t=0.41	s_2	1.01	1.11	1.21	1.31	1.41
s_1						
0.51		0.28	0.37	-0.13	0.31	-0.57
0.61		0.33	-0.10	-0.38	-0.11	0.06
0.71		-0.07	0.54	0.19	0.00	0.53
0.81		0.20	0.18	0.25	0.23	-0.02
0.91		0.33	0.51	-0.07	0.13	**0.66**

t=0.58	s_2	1.18	1.28	1.38	1.48	.158
s_1						
0.68		-0.32	-0.13	0.08	-0.43	-0.14
0.78		-0.01	-0.02	0.05	0.02	0.16
0.88		0.12	0.36	0.45	-0.26	-0.06
0.98		0.30	0.15	-0.50	0.49	**0.60**
1.08		-0.00	-0.29	0.15	0.57	-0.34

t=0.79	s_2	1.39	1.49	1.59	1.69	1.79
s_1						
0.89		-0.21	-0.52	-0.53	-0.11	-0.25
0.99		-0.23	-0.41	-0.21	-0.69	-0.49
1.09		-0.41	0.04	-0.06	0.30	-0.53
1.19		-0.09	-0.09	0.25	-0.18	-0.60
1.29		0.10	-0.12	-1.01	**0.42**	-0.42

t=1.14	s_2	1.74	1.84	1.94	2.04	2.14
s_1						
1.24		**0.36**	0.12	-0.29	-0.48	-0.66
1.34		-0.23	-0.39	-0.38	-0.60	-0.33
1.44		-0.54	-0.09	-0.59	-0.35	0.26
1.54		-0.78	-0.23	-0.51	-0.38	-0.68
1.64		-0.20	-0.03	-0.58	-0.64	-0.24

t=1.42	s_2	2.23	2.33	2.42	2.52	1.61
s_1						
1.75		-0.17	-0.15	-0.28	-0.01	-0.08
1.85		0.22	-0.01	0.06	-0.35	-0.27
1.94		-0.16	-0.20	0.16	-0.12	-0.18
2.04		-0.86	-0.10	-0.06	-0.13	0.04
2.13		0.05	**0.26**	-0.11	-0.37	0.14

t=1.89	s_2	2.16	2.21	2.25	2.30	2.34
s_1						
1.93		-0.00	0.05	0.17	0.07	-0.50
1.98		-0.02	-0.12	-0.34	-0.21	-0.15
2.02		-0.17	-0.23	-0.11	-0.03	0.11
2.07		-0.43	-0.27	0.14	-0.09	-0.16
2.12		-0.07	0.08	**0.29**	0.05	0.13

t=2.15	s_2	2.30	2.33	2.35	2.38	2.40
s_1						
2.18		0.00	-0.06	-0.08	-0.08	0.22
2.20		**0.36**	0.31	0.01	0.08	0.04
2.23		0.03	0.13	0.10	0.28	-0.02
2.25		0.31	0.21	0.04	-0.02	0.25
2.28		0.10	0.26	-0.03	0.14	0.01