Thesis mid-term report

By: Georgy Gomon
Studentnumber: s1559370
Date: November 25, 2021

Description of Site

The thesis is conducted at the medical statistics section of the LUMC. Here I have a desk in a shared office at my disposal. I try to work at the LUMC as much as possible, subject to the changing corona-measures. Weekly meetings take place with my 2 supervisors, R.Tsonaka and B.Mertens. Apart from these weekly meetings I also join other events taking place at the section of medical statistics, such as the weekly lunch-meetings and the book-club. During the weekly lunch-meetings the current research at the medical statistics section is discussed, while in the book club, organized by the PhD students of the section, our statistical knowledge is extended by working through some of the distinguished books in statistics. Currently we are reviewing a book by Bradley Effron about Empirical Bayes Methods [Efron, 2011].

Description of Activities performed so far

The activities performed so far can be divided into 5 main clusters:

- Understanding Joint Mixed Models. For this I needed to refresh my understanding of the course 'Mixed and Longitudinal Modelling' and delve deeper into Joint Models and endogenous covariates. Endogenous covariates were discussed in [P.J. Diggle, 2016], while several examples on how to handle joint models were given in [Fitzmaurice,]. To implement joint models with a scaled linear predictor I took inspiration from joint survival models, see [Rizopoulos, 2012].
- Understanding INLA. To have a better understanding of INLA I started with the original INLA paper, [Rue et al., 2009]. With the help of some other sources I managed to get a basic understanding of what INLA entails. Then I examined the current implementations of INLA on multiple likelihood models and Joint survival models, see [van Niekerk et al., 2021], [van Niekerk et al., 2019] and [Van Niekerk et al., 2019].
- After having developed an understanding of both INLA and Joint Mixed Models I carried on with incorporating the mixed models within the INLA framework. I hereby examined 3 types of models, discussed further in this report (see section 4). Because every model is represented as an Latent Gaussian Model within INLA internally, there are many approaches to implement models within INLA and INLA offers many different 'tricks' to fit a multitude of models. As the implementation of Joint Longitudinal Models is not well documented within INLA, a major part of my activities so far involved finding the different tricks to fit the Joint Models within the INLA framework. In section 4 I mention the methods via which the different models are implemented within INLA.
- To check the validity of the model implementation in INLA I first compared the results obtained with INLA to results obtained using other R-packages, including 'nlme', 'lmer' and 'MCMCglmm'. In this way I could check the consistency of my results as well as discover what the uses and limitations are of these other packages in fitting Joint Models. An overview of the uses and limitations of R-packages in fitting Joint Models is given in section 4.6.
- Following the implementation of all Joint Models in INLA we looked at the Goodness of Fit measures available within INLA and expanded them with some of our own measures. To see the credibility of these Goodness of Fit measures we conducted a simulation study. For this simulation study we simulated data according to each of the models. Each simulated dataset is then fit by all models. In this way we tried to discover whether the models give comparable results and whether a certain type of model always outperforms the others.

Planning of remaining part of project

Having largely completed the set-up of all Joint models within the INLA framework the next step is to apply the methods on a real dataset. For this we shall be using the LUMC Covid dataset. This is a longitudinal dataset of COVID patients treated within the LUMC. Instead of using the entire dataset we shall be focusing on only a few covariates. Furthermore, because of privacy concerns random error has been added to the dataset. The planning of the remaining part of the project will thus mostly be focused on applying the Joint Models to this LUMC dataset. Also, time needs to be reserved for writing the thesis, as the analysis on the LUMC dataset needs to be written and the thesis should be given a coherent structure. The planning of the remaining months can be seen in figure 1.

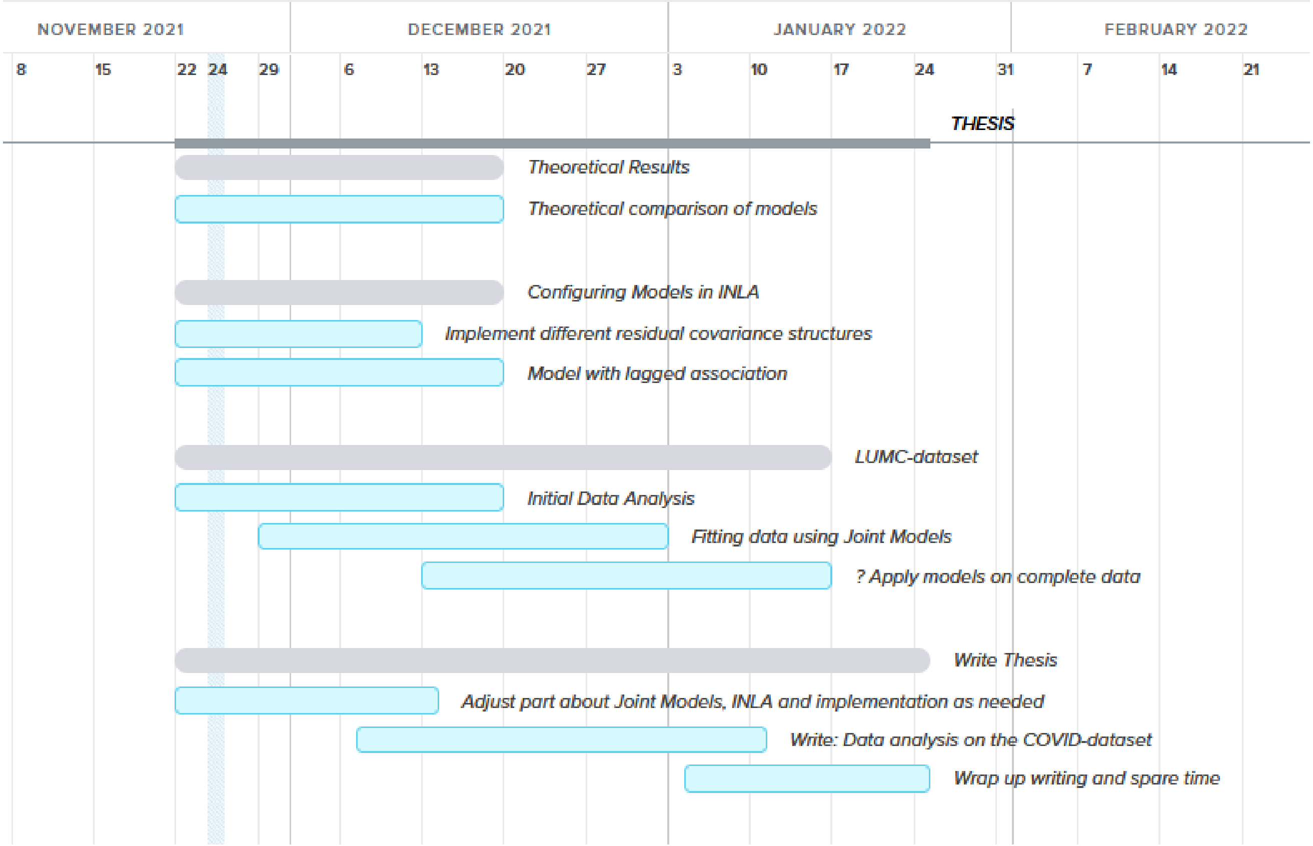

Figure 1: Planning of the remaining part of the project.

Thesis: Progress so far

In the following one can see the current state of the Thesis and the progress made so far.

Abstract

In longitudinal data analysis one often encounters endogenous time-dependent covariates: these are covariates whose current value, given their own history, depends on past values of the outcome. In the presence of such endogenous covariates, because of the cross-reliance of the endogenous covariate on the outcome, standard Mixed Models are no longer valid and one needs to resort to joint modelling of both the outcome and the endogenous covariate. In this thesis several such joint longitudinal models will be discussed. To fit these models we shall be examining a novel Bayesian technique called INLA (Integrated Nested Laplace Integration), which is an elegant technique that could possibly replace the complex and long MCMC estimation procedure. Although INLA has seen rapid development over recent years, joint longitudinal models have so far received little attention. The goal of this thesis is to implement several joint longitudinal models within the INLA framework and apply them to the LUMC Covid dataset.

Contents

1 Introduction **4**

2 Joint Mixed Models **4**
2.1 Mixed Models . . . 4
2.2 Endogenous vs Exogenous covariates . . . 5
2.3 Joint Mixed Models . . . 6
2.3.1 Multivariate Joint Model . . . 6
2.3.2 Joint Mixed Model . . . 6
2.3.3 Mixed Model with scaled linear predictor . . . 6

3 INLA **7**
3.1 Bayesian Inference using INLA . . . 7
3.1.1 Latent Gaussian Model . . . 7
3.1.2 Gaussian Markov Random Field (GMRF) . . . 7
3.1.3 Laplace Approximation . . . 7
3.1.4 Approximating the Latent Field . . . 8
3.1.5 Approximating $p(\theta|\mathbf{y})$ . . . 8
3.1.6 Approximating $p(x_i|\theta, \mathbf{y})$ . . . 8
3.2 Priors in INLA . . . 9
3.2.1 Random Effect Priors . . . 9
3.3 Model Assessment in INLA . . . 9
3.3.1 Marginal Likelihood . . . 9
3.3.2 Conditional Predictive Ordinates (CPO) . . . 10
3.3.3 Probability Integral Transform (PIT) . . . 10
3.3.4 DIC and WAIC . . . 10
3.3.5 Mean Squared Error (MSE) . . . 10

4 Configuring the Joint Models in R-INLA **11**
4.1 Independent Mixed Models . . . 11
4.2 Multivariate Joint Models . . . 12
4.3 Joint Mixed Models . . . 13
4.4 Joint Mixed Models with Scaled Linear Predictor . . . 14
4.5 Summary of Models introduced . . . 16
4.6 Joint Models and their implementation in different R packages . . . 16

5 Simulation Study **16**
5.1 Set-up of Simulation Study . . . 17
5.2 Results of Simulation Study . . . 17

1 Introduction

Longitudinal data analysis focuses on the effect covariates have on a certain outcome over time. As an example we can imagine studying the effect of the covariates ‘sex’, ‘age’ and ‘treatment regime’ on the outcome ‘lung capacity’ following a COVID infection. Within the longitudinal framework we would then measure the values of the covariates and outcome multiple times over the span of e.g. a few months. Within the context of longitudinal data we can split the covariates into 3 groups. Covariates can be time-dependent or time-independent. In our hypothetical example ‘sex’ is a time-independent covariate, as it does not change over time. The time dependent covariates can be split into endogenous and exogenous time-dependent covariates. An exogenous time-dependent covariate is a covariate whose current value, given its own history, does not depend on the value of the outcome at previous measurement times. In our example ‘age’ is such an exogenous time-dependent covariate. ‘Age’ does change over time, but it is independent of ‘lung capacity’ (the outcome) at previous time points. Lastly, we have the endogenous time-dependent covariates, which are covariates whose current value does depend on previous values of the outcome, given their own history. In our example ‘treatment regimen’ is such an endogenous time-dependent covariate, since the lung capacity at previous measurements can influence the treatment regimen the patient is currently receiving, e.g: If the patient is recovering the treatment can be scaled down. Modelling such endogenous time-dependent covariates (we shall call them endogenous covariates) is difficult, since there is a causal path from outcome to endogenous covariate and vice versa. A standard linear mixed model is no longer applicable but instead the endogenous covariate and the outcome need to be modelled jointly. This leads us into the framework of joint longitudinal models. For more information on endogenous covariates and joint models we refer to [P.J. Diggle, 2016]. Within the scope of this thesis the different approaches to joint modelling of the endogenous covariate and the outcome will be studied. The emphasis will be on 3 methods:

- First is a simple multivariate model in which the multiple outcomes are jointly Gaussian distributed. The association between the two outcomes is then modelled via correlated errors in the linear predictors of the outcomes.
- Second is a joint mixed model in which the association between the multiple outcomes is given by multivariate normally distributed random effects and multivariate normally distributed errors terms.
- Lastly a joint model is proposed in which the linear predictor of the endogenous covariate is inserted into the linear predictor of the outcome with an associated scaling factor.

During the Thesis these methods will be applied in R within the Bayesian framework. The emphasis will be to implement the methods using INLA (Integrated Nested Laplace Approximation) and its associated R package R-INLA. INLA is a new Bayesian framework based on Laplace Integration that removes the need for extensive MCMC estimation and is therefore much quicker than standard Bayesian methods. For more information on INLA we refer to [Rue et al., 2009]. For more information about the current joint models implementations of INLA we refer to [van Niekerk et al., 2021].

2 Joint Mixed Models

In this section we shall briefly introduce Mixed Models, before continuing with the definitions of endogenous and exogenous covariates. We shall end this section with the different types of Joint Models used to model endogenous covariates.

2.1 Mixed Models

First we will introduce Mixed Models and the notation that shall be used throughout the Thesis. We shall not go into details about Mixed Models, for those interested we refer to [Fitzmaurice,] and [P.J. Diggle, 2016]. The mixed models we shall be using are of the following form:

$$y_i(t_{ij}) = \mathbf{w}_i \cdot \boldsymbol{\alpha} + \mathbf{v}_i(t_{ij}) \cdot \boldsymbol{\beta} + \mathbf{z}_i(t_{ij}) \cdot \mathbf{b_i} + \epsilon_i(t_{ij})$$

with:

- $y_i(t_{ij})$: Outcome for patient i at time t_{ij}. In total there are $i = 1, ..., N$ patients, with every patient having $j = 1, ..., n_i$ measurements.

- $\mathbf{w}_i$: Vector of fixed time-independent covariates
- $\mathbf{v}_i(t_{ij})$: Vector of fixed time-varying covariates at time t_{ij}.
- $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$: Coefficient vectors of fixed time-independent and time-varying covariates respectively
- $\mathbf{z}_i(t_{ij})$: Vector of random covariates
- $\mathbf{b_i} \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$: Random effects vector
- $\epsilon_i(t_{ij}) \sim \mathcal{N}(0, \sigma^2)$: Residual errors, with $\epsilon_i(t_{ij}) \perp\!\!\!\perp \mathbf{b_i}$ and $\mathbf{b_i} \perp\!\!\!\perp \mathbf{v}_i(t_{ij}), \mathbf{w}_i$.

2.2 Endogenous vs Exogenous covariates

Within longitudinal studies we have both time-invariant and time-varying covariates. Examples of time-invariant covariates are sex, treatment group and genetic profile. Examples of time-varying covariates are age, biomarkers, air-pollution exposure and treatment dose. The time-varying covariates can furthermore be divided into 2 groups: Exogenous and Endogenous covariates. To define exogenous and endogenous covariates the following notation is introduced:

- $y_i(t)$: Value of the response y for subject i at time t.
- $x_i(t)$: Value of the covariate x for subject i at time t.
- $\mathcal{H}_i^Y$: History of the response process of subject i until time t:

$$\mathcal{H}_i^Y(t) = \{y_i(t_{i1}), y_i(t_{i2}), ..., y_i(t_{ik}); t_{ik} < t\}$$

- $\mathcal{H}_i^X$: History of the covariate process of subject i until time t:

$$\mathcal{H}_i^X(t) = \{x_i(t_{i1}), x_i(t_{i2}), ..., x_i(t_{ik}); t_{ik} < t\}$$

- $\mathbf{W}_i$: Vector of time-independent covariates.

Definition 2.1 (Exogenous Covariate) *$X_i(t)$ is an exogenous covariate with respect to the outcome process if the exposure at time t is conditionally independent on the history of the outcome process at time t, given the history of the exposure process at time t. Mathematically,*

$$f\left(x_i(t) | \mathcal{H}_i^Y(t), \mathcal{H}_i^X(t-1), \mathbf{W}_i\right) = f\left(x_i(t) | \mathcal{H}_i^X(t-1), \mathbf{W}_i\right)$$

Thus, for an exogenous covariate the exposure at time t does not depend on previous values of the response. Examples of exogenous covariates are age and air-pollution exposure.
For exogenous covariates the likelihood $f(\mathbf{Y_i}, \mathbf{X_i} | \mathbf{W_i}, \theta)$ can be factorized:

$$\begin{aligned} f(\mathbf{Y_i}, \mathbf{X_i} | \mathbf{W_i}, \theta) &= \left[\prod_{t=1}^{T} f\left(y_i(t) | \mathcal{H}_i^Y(t-1), \mathcal{H}_i^X(t), \mathbf{W}_i, \theta\right)\right] \cdot \left[\prod_{t=1}^{T} f\left(x_i(t) | \mathcal{H}_i^X(t-1), \mathbf{W}_i, \theta\right)\right] = \\ &= \mathcal{L}_Y(\theta_1) \cdot \mathcal{L}_X(\theta_2) \end{aligned} \tag{1}$$

The factorization of the joint likelihood means that we do not need to model the covariate process of X in order to make inference about θ_1 and the outcome Y.

Definition 2.2 (Endogenous Covariate) *$X_i(t)$ is an endogenous covariate with respect to the outcome process if the exposure at time t is conditionally dependent on the history of the outcome process at time t, given the history of the exposure process at time t. Mathematically,*

$$f\left(x_i(t) | \mathcal{H}_i^Y(t), \mathcal{H}_i^X(t-1), \mathbf{W}_i\right) \neq f\left(x_i(t) | \mathcal{H}_i^X(t-1), \mathbf{W}_i\right)$$

Thus, for an exogenous covariate the exposure at time t does depend on previous values of the response. An example might occur in case of a non-controlled study investigating the effect of a certain treatment regimen on symptom severity. If no symptoms are present, the treatment regimen might be made less stringent and vice-versa.
For exogenous covariates the factorization (see equation 1) can not be done, and thus the joint process of X and Y needs to be modelled in order to make inference about Y.

2.3 Joint Mixed Models

As was shown in definition 2.2, in case of endogenous covariates the likelihood $f\left(\mathbf{Y_i}, \mathbf{X_i}|\mathbf{W_i}, \theta\right)$ of the outcome Y and endogenous covariate X can not be factorized and thus both the endogenous covariate X and the outcome Y need to be modelled jointly to make inference.
In this thesis we shall be looking at 3 main methods which enable joint modelling of both the outcome and the endogenous covariate.

2.3.1 Multivariate Joint Model

The first joint model we shall be examining is a multivariate normal joint model. Here the association between the outcome Y and the endogenous covariate X is realized via the residual errors covariance matrix $\mathbf{\Sigma}_i$.

$$\begin{cases} y_i(t_{ij}) = \mathbf{w}_i \cdot \boldsymbol{\alpha} + \mathbf{v}_{yi}^{\mathsf{T}}(t_{ij})\beta_{\mathbf{y}} + \epsilon_{yi}(t_{ij}) \\ x_i(t_{ij}) = \mathbf{w}_i \cdot \boldsymbol{\alpha} + \mathbf{v}_{xi}^{\mathsf{T}}(t_{ij})\beta_{\mathbf{x}} + \epsilon_{xi}(t_{ij}) \end{cases} \quad \text{with} \quad \begin{bmatrix} \epsilon_{yi} \\ \epsilon_{xi} \end{bmatrix} \sim \mathcal{N}_{n_i}(\mathbf{0}, \mathbf{\Sigma}_i)$$

In this model the association can be measured between any pair of time-points and missing data in the response or covariates can be handled simultaneously.
The largest disadvantage of this method is that it allows only for balanced designs and that all covariates and outcomes must be measured at the same time point.
Additionally, a choice must be made for the structure of the variance-covariance matrix $\mathbf{\Sigma}_i$. Possible choices are an Unstructured form (requiring a multitude of parameters to be estimated), Compound symmetry, Autoregressive and Toeplitz.

2.3.2 Joint Mixed Model

The next type of models we shall be examining are joint mixed models of the form shown below:

$$\begin{cases} y_i(t_{ij}) = \mathbf{w}_i \cdot \boldsymbol{\alpha} + \mathbf{v}_{yi}^{\mathsf{T}}(t_{ij})\beta_{\mathbf{y}} + \mathbf{z}_{yi}^{\mathsf{T}}(t_{ij})\mathbf{b}_{yi} + \epsilon_{yi}(t_{ij}) \\ x_i(t_{ij}) = \mathbf{w}_i \cdot \boldsymbol{\alpha} + \mathbf{v}_{xi}^{\mathsf{T}}(t_{ij})\beta_{\mathbf{x}} + \mathbf{z}_{xi}^{\mathsf{T}}(t_{ij})\mathbf{b}_{xi} + \epsilon_{xi}(t_{ij}) \end{cases} \quad \text{with}$$

$$\begin{bmatrix} \mathbf{b}_{yi} \\ \mathbf{b}_{xi} \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}); \qquad \begin{bmatrix} \epsilon_{yi} \\ \epsilon_{xi} \end{bmatrix} \sim \mathcal{N}_{n_i}(\mathbf{0}, \mathbf{\Sigma}_i); \qquad \epsilon_{yi}(t_{ij}) \perp\!\!\!\perp \mathbf{b}_{yi}, \epsilon_{xi}(t_{ij}) \perp\!\!\!\perp \mathbf{b}_{xi}$$

Here association is measured via the random effects (given by the covariance matrix $\mathbf{D}$) and the residual errors (given by covariance matrix $\mathbf{\Sigma}_i$). A large advantage of this model over the Multivariate Joint Model is that here the outcome and endogenous covariate do not need to be measured at the same time, given that one does not incorporate association via the residuals errors in $\mathbf{\Sigma}_i$, thus setting $\mathbf{\Sigma}_i = \mathcal{I}$.

2.3.3 Mixed Model with scaled linear predictor

Lastly we have a mixed model in which the linear predictor of the endogenous covariate is copied into the linear predictor of the outcome with an associated scaling factor γ. Models of this type are very common in Survival analysis, where a longitudinal model and a survival model are combined via a scaling factor. More information about the use of such models within Survival analysis can be found in [Rizopoulos, 2012].
The model is of the following form:

$$\begin{cases} x_i(t_{ij}) = m_i(t_{ij}) + \epsilon_{xi}(t_{ij}) \\ y_i(t_{ij}) = \mathbf{w}_{yi}^{\mathsf{T}}\alpha_{\mathbf{y}} + \gamma \cdot m_i(t_{ij}) + \mathbf{v}_{yi}^{\mathsf{T}}(t_{ij})\beta_{\mathbf{y}} + \mathbf{z}_{yi}^{\mathsf{T}}(t_{ij})\mathbf{b}_{yi} + \epsilon_{yi}(t_{ij}) \end{cases}$$

with

$$m_i(t_{ij}) = \mathbf{w}_{xi}^{\mathsf{T}}\alpha_{\mathbf{x}} + \mathbf{v}_{xi}^{\mathsf{T}}(t_{ij})\beta_{\mathbf{x}} + \mathbf{z}_{xi}^{\mathsf{T}}(t_{ij})\mathbf{b}_{xi}$$

and

$$\begin{aligned} &\mathbf{b}_{xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_x), \qquad \mathbf{b}_{yi} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_y) \\ &\epsilon_{yi}(t_{ij}) \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma_y^2), \qquad \epsilon_{xi}(t_{ij}) \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma_x^2) \\ &\epsilon_{yi}(t_{ij}) \perp\!\!\!\perp \mathbf{b}_{yi}, \qquad \epsilon_{xi}(t_{ij}) \perp\!\!\!\perp \mathbf{b}_{xi} \end{aligned}$$

In the above example the linear predictor of X, $m_i(t_{ij})$, is copied with scaling factor γ only at time-point t_{ij}. However, this dependence can be elaborated and linear predictors at any time point $t_{ik} < t_{ij}$ can be included into the linear predictor of the outcome, in this way creating a lagged functional form.

3 INLA

In this thesis we shall be fitting the joint models using Integrated Nested Laplace Approximation (INLA) and it's implementation in R with the R-package R-INLA [Rue et al., 2009]. INLA combines the usage of Latent Gaussian Models (LGM's), Gaussian Markov Random Fields (GMRF's), Numerical methods for sparse matrices and Laplace approximations to derive approximate Bayesian inference. Overall INLA is much faster than Monte Carlo Markov Chain (MCMC) methods for Bayesian inference and does not necessitate the need for long sampling chains. Also, research has shown that in terms of accuracy INLA is not inferior to MCMC methods (NEEDS A REFERENCE!!)
First a quick overview of INLA and it's package R-INLA will be given, before continuing with the implementation of the joint models into the INLA framework.

3.1 Bayesian Inference using INLA

3.1.1 Latent Gaussian Model

INLA uses the fact that many statistical models, including the joint longitudinal models discussed in this thesis, can be rewritten as a Latent Gaussian Model (LGM). In addition, only models that can be rewritten as LGM's can be used within the INLA framework.
An LGM consists of the following elements:

- Likelihood of the outcome: $\mathbf{y}|\mathbf{x}, \theta_2 \sim \prod_i p(y_i|\eta_i, \theta_2)$
- The Latent Field: $\mathbf{x}|\theta_1 \sim p(\mathbf{x}, \theta_1) = \mathcal{N}(0, \mathbf{\Sigma})$.
- The Hyperpriors: $\theta = (\theta_1, \theta_2) \sim p(\theta)$.

Here $\mathbf{y}$ is the observed data and $\mathbf{x}$ are all the parameters in the linear predictor. Note that the dimension of $\mathbf{x}$ is usually very large (model has many data-points), but the dimension of θ is usually small (just a few parameters are needed to define the random effects structure).
Imagine we have the most general form of a generalized linear mixed model:

$$y \sim \prod_i^N p(y_i|\mu_i) \quad \text{with} \quad g(\mu_i) \equiv \eta_i = \alpha + \sum_{k=1}^{n_\beta} \beta_k \cdot z_{k_i} + \sum_{j=1}^{n_f} f^{(j)}(w_{ji}) + \epsilon_i$$

Here $g()$ is the link function, α the intercept, β the regression parameters of covariates z and $f()$ the random effects of covariates w.
Such a model is an LGM if and only if we assume that all parameters have a joint Normal distribution, thus:

$$\mathbf{x} = [\eta, \alpha, \beta, f()] \sim \mathcal{N}(0, \Sigma)$$

If we furthermore assume conditional independence in $\mathbf{x}$, then this latent field $\mathbf{x}$ is a Gaussian Markov Random Field.

3.1.2 Gaussian Markov Random Field (GMRF)

Our vector $\mathbf{x} = [\eta, \alpha, \beta, f()] \sim \mathcal{N}(0, \Sigma)$ can now be thought of as a Gaussian Markov Random Field (GMRF). A GMRF is a normally distributed random vector $\mathbf{x} = (x_1, ..., x_n)$ with Markov properties, such as that for some $i \neq j$, $x_i \perp\!\!\!\perp x_j|\mathbf{x}_{-ij}$, which means that x_i is independent of x_j given all elements of $\mathbf{x}$ other than i and j ($\mathbf{x}_{-ij}$). The Markov properties are given in the Precision matrix $Q = \Sigma^{-1}$, which is the inverse of the covariance matrix. Rue et all [Rue et al., 2009] showed that $x_i \perp\!\!\!\perp x_j|\mathbf{x}_{-ij} \quad \text{iff} \quad Q_{ij} = 0$. This result ensures that if in our vector $\mathbf{x} = [\eta, \alpha, \beta, f()] \sim \mathcal{N}(0, \Sigma)$ the different elements are independent, the precision matrix Q will be very sparse, allowing for easy and fast computations.

3.1.3 Laplace Approximation

INLA uses the Laplace Approximation to estimate any distribution $g(x)$ with a normal distribution. The first 3 terms of the Taylor expansion around the mode ($\hat{x}$) are used to approximate log $g(x)$ by:

$$\log g(x) \approx \log g(\hat{x}) + \frac{\delta \log g(\hat{x})}{\delta x}(x - \hat{x}) + \frac{\delta^2 \log g(\hat{x})}{2\delta x^2}(x - \hat{x})^2$$

The second term in this approximation, $\frac{\delta \log g(\hat{x})}{\delta x}(x-\hat{x})$, equals 0, since we are considering the derivative at the mode which is a maximum of the function.
We now estimate the variance as:

$$\hat{\sigma}^2 = -\frac{2\delta x^2}{\delta^2 \log g(\hat{x})}\bigg|_{\hat{x}}$$

Using this we obtain:

$$\log g(x) \approx \log g(\hat{x}) - \frac{1}{2\sigma^2}(x-\hat{x})^2$$

With the last expression we can perform a normal approximation:

$$\int g(x)dx = \int \exp\left[\log g(x)\right]dx \approx \int \exp\left[\log g(\hat{x}) - \frac{1}{2\sigma^2}(x-\hat{x})^2\right]dx =$$

$$= \exp\left[\log g(\hat{x})\right] \cdot \int \exp\left[-\frac{1}{2\sigma^2}(x-\hat{x})^2\right]dx = \text{constant} \cdot \int \exp\left[-\frac{1}{2\sigma^2}(x-\hat{x})^2\right]dx$$

Thus, the distribution of $g(x)$ is now approximated by a normal distribution with mean $\hat{x}$, which is found by solving $g'(x) = 0$ and with variance $\hat{\sigma}^2 = -\frac{2\delta x^2}{\delta^2 \log g(\hat{x})}\Big|_{\hat{x}}$, obtained at the mode $\hat{x}$.

3.1.4 Approximating the Latent Field

When conducing Bayesian inference we are interested in the marginals for the elements of the latent field (e.g: regression coefficients):

$$p(x_i|\mathbf{y}) = \int p(x_i, \theta|\mathbf{y})d\theta = \int p(x_i|\theta, \mathbf{y})p(\theta|\mathbf{y})d\theta$$

and the elements of the hyperprior distribution (e.g: variances of random effects):

$$p(\theta_k|\mathbf{y}) = \int p(\theta|\mathbf{y})d\theta_{-k}$$

To obtain these estimates we need to approximate $p(x_i|\theta, \mathbf{y})$ and $p(\theta|\mathbf{y})$.

3.1.5 Approximating $p(\theta|\mathbf{y})$

We can approximate the marginal distribution as:

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{x}, \theta|\mathbf{y})}{p(\mathbf{x}|\theta, \mathbf{y})} \approx \frac{p(\mathbf{y}|\mathbf{x}, \theta)p(\mathbf{x}|\theta)p(\theta)}{\tilde{p}(\mathbf{x}|\theta, \mathbf{y})}\bigg|_{x=x^*(\theta)} = \tilde{p}(\theta|\mathbf{y}).$$

Here a Gaussian Laplace approximation is used for the denominator $p(\mathbf{x}|\theta, \mathbf{y})$ at the mode $x = x^*(\theta)$.

3.1.6 Approximating $p(x_i|\theta, \mathbf{y})$

To approximate $p(x_i|\theta, \mathbf{y})$ INLA has 3 options:

- Normal approximation, used in INLA when selecting the option 'Gaussian'. Here we approximate $p(x_i|\theta, \mathbf{y})$ using standard Laplace approximation, and since we already computed $\tilde{p}(\mathbf{x}|\theta, \mathbf{y})$ during the exploration of $p(\theta|\mathbf{y})$ only the marginals are left to be computed. This method is by far the fastest of the three but often yields poor results.

- Laplace approximation, used in INLA when selecting the option 'Laplace'. Partitions the latent field $\mathbf{x} = [x_j, \mathbf{x}_{-j}]$ and uses Laplace approximation for each element x_j in the latent field:

  $$p(x_j|\theta, \mathbf{y}) \propto \frac{p(\mathbf{x}, \theta|\mathbf{y})}{p(\mathbf{x}_{-j}|x_j, \theta, \mathbf{y})} \propto \frac{p(\theta)p(\mathbf{x}|\theta)p(\mathbf{y}|\mathbf{x})}{p(\mathbf{x}_{-j}|x_j, \theta, \mathbf{y})}$$

  Overall gives good results because the conditionals $p(\mathbf{x}_{-j}|x_j, \theta, \mathbf{y})$ are often close to normal, but is computationally expensive.

- Simplified Laplace approximation, default setting in INLA. Uses a compromise between the first 2 methods. Is computationally fast and almost always gives results very similar to the Laplace approximation.

For more information regarding Bayesian Inference with INLA we refer to [Rue et al., 2009].

3.2 Priors in INLA

The prior for fixed effects in INLA is a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, in which both the mean and precision $\tau = 1/\sigma^2$ can be specified. The default values supplied by INLA are $\mu = 0, \tau = 0.001$.

3.2.1 Random Effect Priors

Within the scope of this thesis we shall mainly be using independent and identically distributed random effect structures. Suppose u and v are two random effects, and together they have an i.i.d. bivariate Normal distribution:

$$\begin{bmatrix} u \\ v \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{W}^{-1}), \quad \text{with covariance matrix} \quad \mathbf{W}^{-1} = \begin{pmatrix} 1/\tau_u & \rho/\sqrt{\tau_u \tau_v} \\ \rho/\sqrt{\tau_u \tau_v} & 1/\tau_v \end{pmatrix}$$

Here τ_u, τ_v (marginal precisions) and ρ (correlation coefficient) are hyperparameters.
The hyperparameters are represented internally in INLA as $\theta = (\log \tau_u, \log \tau_v, \phi)$, with $\rho = 2\frac{\exp(\phi)}{\exp(\phi)+1} - 1$.
As we are more interested in the variances σ_u^2 and σ_v^2 rather than the precisions τ_u and τ_v we use the inverse of the posterior marginal distribution of the precisions to obtain the corresponding distributions of the variances.
The precision matrix $\mathbf{W}$ is a $p = 2$ dimensional Wishart distribution with support n:

$$\mathbf{W} \sim Wishart_2(n, \mathbf{R}^{-1}) \quad \text{with} \quad \mathbf{R} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \quad \text{and} \quad R_{12} = R_{21} \text{ due to symmetry}$$

Some properties of the Wishart distribution are:

$$\mathbb{E}(\mathbf{W}) = n\mathbf{R}^{-1}, \quad \mathbb{E}(\mathbf{W}^{-1}) = \frac{\mathbf{R}}{n - (p+1)}$$

The variance of the Wishart distribution has no direct overall form, but in general the variance is larger with increasing support n.
The iid random effects thus have prior-parameters $n,\ R_{11},\ R_{21} = R_{12}$ and R_{22}. These can be specified with default values (for the case $p = 2$) being $(4, 1, 1, 0)$.

3.3 Model Assessment in INLA

Several methods are implemented in INLA to assess the goodness of fit of a model.

3.3.1 Marginal Likelihood

The Marginal Likelihood, also called Model evidence, is the probability that the data observed originates from a given model, independent of the parameters of that model (the parameters of the model are integrated out).
The Marginal Likelihood is a very convenient exclusively Bayesian model assessment tool which enables the comparison between models.
In INLA the Marginal Likelihood is approximated as:

$$\widetilde{\pi}(y) = \int \left. \frac{\pi(\theta, x, y)}{\widetilde{\pi}_G(x|\theta, y)} \right|_{x=x^*(\theta)} d\theta$$

Here $\widetilde{\pi}_G(x|\theta, y)$ is the Gaussian approximation (see section 3.1.3) at the mode $x = x^*(\theta)$.
When considering a set of M models $\{\mathcal{M}_m\}_{m=1}^M$, the marginal likelihoods are written down as $\pi(y|\mathcal{M}_m)$. If supplying each model with a prior $\pi(\mathcal{M}_m)$, posterior probabilities for each of the models can be calculated as: $\pi(\mathcal{M}_m|y) \propto \pi(y|\mathcal{M}_m)\pi(\mathcal{M}_m)$. These posteriors can now be used to compute the Bayes factor K for 2 different models $\mathcal{M}_1$ and $\mathcal{M}_2$:

$$K = \frac{\pi(\mathcal{M}_1|y)}{\pi(\mathcal{M}_2|y)} = \frac{\pi(y|\mathcal{M}_1)\pi(\mathcal{M}_1)}{\pi(y|\mathcal{M}_2)\pi(\mathcal{M}_2)}$$

In case of equal priors for the 2 models (the models are considered equally likely), the Bayes factor is simply the fraction of the Marginal Likelihoods of the models:

$$K = \frac{\pi(\mathcal{M}_1|y)}{\pi(\mathcal{M}_2|y)} = \frac{\pi(y|\mathcal{M}_1)}{\pi(y|\mathcal{M}_2)}.$$

The strength of evidence of a Bayes Factor $K < 3.2$ is considered very weak, while only a Bayes Factor of $K > 10$ is considered to be indicative of strong evidence of model $\mathcal{M}_1$ versus model $\mathcal{M}_2$. INLA works with the natural logarithm of the Bayes Factor K, meaning that a difference in logarithms $\ln(\mathcal{M}_1) - \ln(\mathcal{M}_2) < 1.16$ indicates very weak evidence while a different of $\ln(\mathcal{M}_1) - \ln(\mathcal{M}_2) > 2.3$ indicates strong evidence.

3.3.2 Conditional Predictive Ordinates (CPO)

The Conditional Predictive Ordinate (CPO) is computed for each observation i as:

$$CPO_i = \pi(y_i|y_{-i}).$$

It is the posterior probability of observing observation y_i when the model is fit using all data but y_i. A small value for an observation might indicate a possible outlier. INLA approximates this quantity for every observation without the need the re-analyse the model with the given observation removed.
The CPO can be summarized over all the data by:

$$CPO = -\sum_{i=1}^{N} \log(CPO_i)$$

A smaller value indicates a better fit of the model over all observations.

3.3.3 Probability Integral Transform (PIT)

The Probability Integral Transform (PIT) is very similar to the CPO and is computed for each observation as:

$$PIT_i = \pi(y_i^{new} \leq y_i|y_{-i})$$

The PIT measures the probability for a new observation y_i^{new} to be lower than the actual observation y_i when the model is fit using all data but y_i. Both the CPO and PIT thus apply techniques very similar to Leave-One-Out Cross-Validation (LOO CV). A very large or small PIT for a given value indicates a possibly surprising observation.
Over all the observations, in case of a good model, the PIT's should be approximately uniformly distributed on $[0, 1]$. The Kolmogorov Smirnov non-parametric test is used to test whether the PIT's are indeed uniformly distributed.

3.3.4 DIC and WAIC

The DIC (Deviance Information Criteria) is a popular method for model selection, as it combines goodness of fit with penalization of the number of parameters used by the model. The DIC is given by:

$$DIC = D(\hat{x}, \hat{\theta}) + 2p_D$$

Here $D(\hat{x}, \hat{\theta})$ is the model deviance, which is calculated using the posterior mean $\hat{x}$ and the posterior mode $\hat{\theta}$, as the distribution of θ can be severely skewed.
The effective number of parameters p_D is approximated as:

$$p_D(\theta) \approx n - tr\{Q(\theta)Q*(\theta)^{-1}\}$$

With n being the number of observations and Q being the precision of the Gaussian Markov Random Field, see section 3.1.2.
The Watanabe-Akaike Information Criterion is similar to the DIC, with the only difference being that the effective number of parameters p_D is calculated in a different way.

3.3.5 Mean Squared Error (MSE)

The last method via which we shall assess goodness of fit is Mean Squared Error (MSE). MSE is not incorporated into the INLA package but was calculated using the posterior means of fitted values. The MSE is given by

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Here N is the total number of measurements while y_i and $\hat{y}_i$ are the actual and fitted (posterior means) outcomes respectively. In order to asses both the marginal and hierarchical model fit properties 3 types of MSE were calculated:

- MSE_{train}: Here the MSE is calculated based on the training set, thus using the data that was used to fit the model. It is useful in order to test how well the model can fit the training subjects and to discover possible instances of overfitting.
- MSE_{same}: MSE determined on subsequent observations of subjects in the training set. These are thus subjects whose random effects have been determined by the model, thus giving hierarchical results. In this way one can inspect how well the model is able to fit the random effects of each individual.
- MSE_{test}: MSE calculated based on test subjects, with the random effects of each subject unknown. Here the interest is only on the marginal results, thus solely showing the ability of the model to correctly estimate fixed effects.

4 Configuring the Joint Models in R-INLA

The main part of the thesis consisted of configuring the Joint Models within the R-INLA package and testing them on simulated data. During this testing phase the results obtained using R-INLA were compared with results obtained using the R-packages nlme, lmer and MCMCglmm.

4.1 Independent Mixed Models

The first model we used to test the implementation of Joint Mixed Models in R-INLA was a simple Mixed Model without any association between the outcomes y and the endogenous covariate x. Throughout the thesis this shall be seen as a baseline model to compare the other models to, as this model does not capture the association between the endogenous covariate and the outcome, and thus would theoretically not be able to fit well in the presence of a endogenous covariate.

Definition 4.1 (Model 0) *We shall be referring to this model, without any association between the endogenous covariate and the outcome, as Model 0. Here we have 2 independent mixed models, one for the endogenous covariate x and one for the outcome y. In each mixed model a random intercept and random slope are introduced. The Conditional Independence Assumption is presumed to be true, the assumption that the random effects capture all correlation between measurements on a patient, and thus no correlation is left for the errors. Mathematically, the model is specified as:*

$$\begin{cases} x_{i,j} = (\beta_0^{(x)} + u_{0,i}^{(x)}) + \beta_v^{(x)} \cdot v_{i,j} + \left(\beta_t^{(x)} + u_{t,i}^{(x)}\right) \cdot t_{i,j} + \epsilon_{i,j}^{(x)} \\ y_{i,j} = (\beta_0^{(y)} + u_{0,i}^{(y)}) + \beta_v^{(y)} \cdot v_{i,j} + \left(\beta_t^{(y)} + u_{t,i}^{(y)}\right) \cdot t_{i,j} + \epsilon_{i,j}^{(y)} \end{cases} \quad \textit{with}$$

$$\begin{bmatrix} u_{0,i}^{(x)} \\ u_{t,i}^{(x)} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma_{x,0}^2 & \sigma_{x,(0,t)} \\ \sigma_{x,(t,0)} & \sigma_{x,t}^2 \end{pmatrix} \right] ; \quad \begin{bmatrix} u_{0,i}^{(y)} \\ u_{t,i}^{(y)} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma_{y,0}^2 & \sigma_{y,(0,t)} \\ \sigma_{y,(t,0)} & \sigma_{y,t}^2 \end{pmatrix} \right]$$

$$\begin{bmatrix} \boldsymbol{\epsilon}_i^{(x)} \\ \boldsymbol{\epsilon}_i^{(y)} \end{bmatrix} \sim \mathcal{N}_{2j}(\mathbf{0}, \mathbf{I}_{2j})$$

Here we use the following notation (which will also be used for the remainder of this chapter):

- *$x_{i,j}$ and $y_{i,j}$: The endogenous covariate and outcome respectively for patients $i = 1, ..., N$ at time-points $j = 1, ..., n_i$.*
- *$\beta_0^{(x)}$, $\beta_v^{(x)}$ & $\beta_t^{(x)}$: The fixed effects for the intercept, covariate $v_{i,j}$ and time $t_{i,j}$ (taken to be linear) for the endogenous covariate x. $\beta_0^{(y)}$, $\beta_v^{(y)}$ & $\beta_t^{(y)}$ have similar roles for the outcome y.*
- *$u_{0,i}^{(x)}$ & $u_{t,i}^{(x)}$: Random intercept and random (time)-slope for patient i for endogenous covariate x. Together they are normally distributed with mean 0 and covariance matrix $\begin{pmatrix} \sigma_{x,0}^2 & \sigma_{x,(0,t)} \\ \sigma_{x,(t,0)} & \sigma_{x,t}^2 \end{pmatrix}$.*
  $u_{0,i}^{(y)}$ & $u_{t,i}^{(y)}$: similarly the Random Intercept and Random slope for the outcome y.
- *$\epsilon_{i,j}^{(x)}$ & $\epsilon_{i,j}^{(y)}$ are the errors for the endogenous covariate x and the outcome y respectively.*

To fit this multiple-likelihood model within the R-INLA framework a few tricks have to be used:

- The response variables $\mathbf{x}$ and $\mathbf{y}$, as well as the fixed effects (Intercept, covariate $\mathbf{v}$ as well as time $\mathbf{t}$) have to be stored in matrices of the following form:

$$\begin{array}{cccc}
\text{response variables} & & & \\
\mathbf{y}^1 \text{ and } \mathbf{y}^2 & \text{Intercept} & \text{Covariate } \mathbf{v} & \text{time } \mathbf{t} \\
\begin{bmatrix} x_{1,1} & NA \\ x_{1,2} & NA \\ \vdots & \vdots \\ x_{N,n_N} & NA \\ NA & y_{1,1} \\ NA & y_{1,2} \\ \vdots & \vdots \\ NA & y_{N,n_N} \end{bmatrix}, &
\begin{bmatrix} 1 & NA \\ 1 & NA \\ \vdots & \vdots \\ 1 & NA \\ NA & 1 \\ NA & 1 \\ \vdots & \vdots \\ NA & 1 \end{bmatrix}, &
\begin{bmatrix} v_{1,1} & NA \\ v_{1,2} & NA \\ \vdots & \vdots \\ v_{N,n_N} & NA \\ NA & v_{1,1} \\ NA & v_{1,2} \\ \vdots & \vdots \\ NA & v_{N,n_N} \end{bmatrix}, &
\begin{bmatrix} t_{1,1} & NA \\ t_{1,2} & NA \\ \vdots & \vdots \\ t_{N,n_N} & NA \\ NA & t_{1,1} \\ NA & t_{1,2} \\ \vdots & \vdots \\ NA & t_{N,n_N} \end{bmatrix}
\end{array}$$

  This ensures that INLA can fit two likelihoods and that INLA knows which covariates belong to which Likelihood.

- In order to fit the dependent random slope and random intercept, one has to use the 'iid2d' correlated random effect structure in INLA together with the 'copy' feature.
  An example of such a command in INLA is:

```
f(Intercept, model="iid2d", n=2*N)+f(Time, time, copy="Intercept")
```

  Using this tells INLA that we shall be using a iid2d random effect structure (described in section 3.2.1) with a total of $2N$ random effects (2 for each individual, 1 random intercept + 1 random slope). The copy feature tells INLA that the 'Time' term is the second of these dependent random effects. One should also tell INLA which random effects are shared by the same object. If there are 2 measurements per person, the random effects for Intercept and Time should thus be written down as:

$$\begin{array}{cc}
\text{Random Intercept} & \text{Random Slope} \\
\begin{bmatrix} 1 & NA \\ 1 & NA \\ 2 & NA \\ \vdots & \vdots \\ N & NA \\ NA & 1 \\ NA & 1 \\ NA & 2 \\ \vdots & \vdots \\ NA & N \end{bmatrix}, &
\begin{bmatrix} 1 & NA \\ 1 & NA \\ 2 & NA \\ \vdots & \vdots \\ N & NA \\ NA & 1 \\ NA & 1 \\ NA & 2 \\ \vdots & \vdots \\ NA & N \end{bmatrix}
\end{array}$$

  This ensures that the random effects are unique per subject and in total we have $2 \cdot N$ different random effects.

The tricks presented in this section will be used to fit all subsequent models in this thesis. New R-INLA features will be discussed as they are used.
Model 0 can also be fit using the R-packages NLME, LMER and MCMCglmm. The details will not be presented here, but for those interested we refer to the open github repository where all code used during the thesis can be found `https://github.com/georgygomon/Thesis_open`.

4.2 Multivariate Joint Models

Having completed a base-line model which does not account for the association between endogenous covariates and the outcome we shall look at the first type of joint model which does take into account this association, the Multivariate Joint Model (see section 2.3.1).

Definition 4.2 (Model 1A) *This Model, in which the association between the endogenous covariate x and* *the outcome y is modelled via residual errors, we shall be referring to as Model 1A.*

$$\begin{cases} x_{i,j} = \beta_0^{(x)} + \beta_v^{(x)} \cdot v_i + \beta_t^{(x)} \cdot t_{i,j} + \epsilon_{i,j}^{(x)} \\ y_{i,j} = \beta_0^{(y)} + \beta_v^{(y)} \cdot v_i + \beta_t^{(y)} \cdot t_{i,j} + \epsilon_{i,j}^{(y)} \end{cases} \quad \textit{with} \quad \begin{bmatrix} \epsilon_{i,1}^{(x)} \\ \epsilon_{i,2}^{(x)} \\ \epsilon_{i,1}^{(y)} \\ \epsilon_{i,2}^{(y)} \end{bmatrix} \sim \mathcal{N}_4 \left[\mathbf{0}, \begin{pmatrix} \sigma_{x,1}^2 & \cdots & \cdots & \cdots \\ \cdots & \sigma_{x,2}^2 & \cdots & \cdots \\ \cdots & \cdots & \sigma_{y,1}^2 & \cdots \\ \cdots & \cdots & \cdots & \sigma_{y,2}^2 \end{pmatrix} \right]$$

To fit this model using R-INLA an additional trick has to be used, as INLA does not allow for dependent residual errors.

- Since INLA does not allow for dependent residual errors, the residual errors should be modelled in INLA as random effects u rather than errors. Simultaneously, to get rid of the Gaussian noise of linear regression, one sets the Gaussian errors fixed with a very high precision τ, in this way eliminating it. The resulting model is thus actually:

  $$\begin{cases} x_{i,j} = \beta_0^{(x)} + \beta_v^{(x)} \cdot v_i + \beta_t^{(x)} \cdot t_{i,j} + u_{i,j}^{(x)} + \epsilon_{i,j}^{(x)} \\ y_{i,j} = \beta_0^{(y)} + \beta_v^{(y)} \cdot v_i + \beta_t^{(y)} \cdot t_{i,j} + u_{i,j}^{(y)} + \epsilon_{i,j}^{(y)} \end{cases}$$

  with

  $$\begin{bmatrix} u_{i,1}^{(x)} \\ u_{i,2}^{(x)} \\ u_{i,1}^{(y)} \\ u_{i,2}^{(y)} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma_{x,1}^2 & \cdots & \cdots & \cdots \\ \cdots & \sigma_{x,2}^2 & \cdots & \cdots \\ \cdots & \cdots & \sigma_{y,1}^2 & \cdots \\ \cdots & \cdots & \cdots & \sigma_{y,2}^2 \end{pmatrix} \right] \quad \text{and} \quad \begin{bmatrix} \boldsymbol{\epsilon}_i^{(x)} \\ \boldsymbol{\epsilon}_i^{(y)} \end{bmatrix} \sim \mathcal{N}_{2j}(\mathbf{0}, \tau \cdot \mathbf{I}_{2j})$$

  The random effects u have an *iid4d* distribution, as was discussed in sections 3.2.1 and 4.1. Since the precision τ is set to be very high, the gaussian noise is thereby practically eliminated and thus this alternative representation in INLA corresponds very well to the actual Model 1A. However, within the R-INLA package one can not have an iid random effect with more than 5 components, and thus this implementation is limited to just 2 observations per subject, as for 3 observations per subject we would need $3 \cdot 2 = 6 > 5$ components.

Model 1A can also be implemented in the R-packages nlme (function gls) and MCMCglmm.

4.3 Joint Mixed Models

We shall now continue with Joint Mixed Models where the association between the endogenous covariate and the outcome is modelled via dependence on the random effects.

Definition 4.3 (Model 2A) *We start with a model in which the random intercept of both the endogenous covariate x and the outcome y are dependent. The association in time is given via the correlated residual errors. This model we shall refer to as model 2A.*

$$\begin{cases} x_{i,j} = (\beta_0^{(x)} + u_{0,i}^{(x)}) + \beta_v^{(x)} \cdot v_i + \beta_t^{(x)} \cdot t_{i,j} + \epsilon_{i,j}^{(x)} \\ y_{i,j} = (\beta_0^{(y)} + u_{0,i}^{(y)}) + \beta_v^{(y)} \cdot v_i + \beta_t^{(y)} \cdot t_{i,j} + \epsilon_{i,j}^{(y)} \end{cases}$$

with

$$\begin{bmatrix} u_{0,i}^{(x)} \\ u_{0,i}^{(y)} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma_{x,0}^2 & \sigma_{(x,y),0} \\ \sigma_{(y,x),0} & \sigma_{y,0}^2 \end{pmatrix} \right]; \quad \begin{bmatrix} \epsilon_{i,1}^{(x)} \\ \epsilon_{i,2}^{(x)} \\ \epsilon_{i,1}^{(y)} \\ \epsilon_{i,2}^{(y)} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma_{x,1}^2 & \cdots & \cdots & \cdots \\ \cdots & \sigma_{x,2}^2 & \cdots & \cdots \\ \cdots & \cdots & \sigma_{y,1}^2 & \cdots \\ \cdots & \cdots & \cdots & \sigma_{y,2}^2 \end{pmatrix} \right]$$

To fit this model in INLA one simply combines the methods shown when discussing the models in Definitions 4.1 and 4.2. Model 2A can also be fit using nlme. A disadvantage of model 2A is that it can only fit data with no more than 2 measurements per subject, because of the limitations of the residual errors covariance structure.

Definition 4.4 (Model 2A1) *Model 2A1 is a small modification from model 2A in which the covariances between the residual errors are set to 0. This allows us to fit the model in case of more than 2 measurements per subject. Note that this model has no random-slope nor does it have any means of modelling a subject-specific change in time. Mathematically, the only difference between models 2A1 and 2A is a different residual error covariance matrix:*

$$\begin{bmatrix} \epsilon_{i,1}^{(x)} \\ \epsilon_{i,2}^{(x)} \\ \epsilon_{i,1}^{(y)} \\ \epsilon_{i,2}^{(y)} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma_{x,1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{x,2}^2 & 0 & 0 \\ 0 & 0 & \sigma_{y,1}^2 & 0 \\ 0 & 0 & 0 & \sigma_{y,2}^2 \end{pmatrix} \right]$$

Definition 4.5 (Model 2B) *Next we consider a model in which both the association at baseline and in time between the endogenous covariate and the outcome are modelled via random effects. The residual errors are independent per measurement and play no role in the association. We shall call this Model 2B.*

$$\begin{cases} x_{i,j} = (\beta_0^{(x)} + u_{0,i}^{(x)}) + \beta_v^{(x)} \cdot v_i + \left(\beta_t^{(x)} + u_{t,i}^{(x)}\right) \cdot t_{i,j} + \epsilon_{i,j}^{(x)} \\ y_{i,j} = (\beta_0^{(y)} + u_{0,i}^{(y)}) + \beta_v^{(y)} \cdot v_i + \left(\beta_t^{(y)} + u_{t,i}^{(y)}\right) \cdot t_{i,j} + \epsilon_{i,j}^{(y)} \end{cases}$$

with

$$\begin{bmatrix} u_{0,i}^{(x)} \\ u_{0,i}^{(y)} \\ u_{t,i}^{(x)} \\ u_{t,i}^{(y)} \end{bmatrix} \sim \mathcal{N}_4 \left[\mathbf{0}, \begin{pmatrix} \sigma_{x,0}^2 & \sigma_{(x,0),(y,0)} & \sigma_{(x,0),(x,t)} & \sigma_{(x,0),(y,t)} \\ \sigma_{(y,0),(x,0)} & \sigma_{y,0}^2 & \sigma_{(y,0),(x,t)} & \sigma_{(y,0),(y,t)} \\ \sigma_{(x,t),(x,0)} & \sigma_{(x,t),(y,0)} & \sigma_{x,t}^2 & \sigma_{(x,t),(y,t)} \\ \sigma_{(y,t),(x,0)} & \sigma_{(y,t),(y,0)} & \sigma_{(y,t),(x,t)} & \sigma_{y,t}^2 \end{pmatrix} \right] ; \quad \begin{bmatrix} \boldsymbol{\epsilon}_i^{(x)} \\ \boldsymbol{\epsilon}_i^{(y)} \end{bmatrix} \sim \mathcal{N}_{2j}(\mathbf{0}, \mathbf{I}_{2j})$$

This model can be fit in INLA using the 'iid4d' random effects (see section 3.2.1) and the tricks shown in definition 4.1. Furthermore, this model can also be fit using NLME, LMER and MCMCglmm.

Definition 4.6 (Model 2B1) *Model 2B1 is a small modification from model 2B and the two models are nested. The only difference is that in model 2B1 there is no dependence between the random slope and random intercept in either x or y. Thus, the covariance matrix of the random effects becomes:*

$$\begin{bmatrix} u_{0,i}^{(x)} \\ u_{0,i}^{(y)} \\ u_{t,i}^{(x)} \\ u_{t,i}^{(y)} \end{bmatrix} \sim \mathcal{N}_4 \left[\mathbf{0}, \begin{pmatrix} \sigma_{x,0}^2 & \sigma_{(x,0),(y,0)} & 0 & 0 \\ \sigma_{(y,0),(x,0)} & \sigma_{y,0}^2 & 0 & 0 \\ 0 & 0 & \sigma_{x,t}^2 & \sigma_{(x,t),(y,t)} \\ 0 & 0 & \sigma_{(y,t),(x,t)} & \sigma_{y,t}^2 \end{pmatrix} \right]$$

This model was introduced to test, based on the goodness of fit measures, how similar these 2 nested models would perform on simulated data.

4.4 Joint Mixed Models with Scaled Linear Predictor

The last type of models we shall be considering in this thesis are models in which the linear predictor for the endogenous covariate x, m_{ij}, is copied with a scaling factor γ into the linear predictor of the outcome y.

Definition 4.7 (Model 3A) *In this model the entire linear predictor of the endogenous covariate x is copied into the linear predictor of the outcome y with scaling factor gamma. In both linear predictors we have dependent random intercept and random slope terms. The model will be referred to as Model 3A and is mathematically given by:*

$$\begin{cases} m_{ij} = (\beta_0^{(x)} + u_{0,i}^{(x)}) + \beta_v^{(x)} \cdot v_i + \left(\beta_t^{(x)} + u_{t,i}^{(x)}\right) \cdot t_{i,j} \\ x_{i,j} = m_{ij} + \epsilon_{i,j}^{(x)} \\ y_{i,j} = \gamma \cdot m_{ij} + (\beta_0^{(y)} + u_{0,i}^{(y)}) + \beta_v^{(y)} \cdot v_i + \left(\beta_t^{(y)} + u_{t,i}^{(y)}\right) \cdot t_{i,j} + \epsilon_{i,j}^{(y)} \end{cases}$$

with

$$\begin{bmatrix} u_{0,i}^{(x)} \\ u_{t,i}^{(x)} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma_{x,0}^2 & \sigma_{x,(0,t)} \\ \sigma_{x,(t,0)} & \sigma_{x,t}^2 \end{pmatrix} \right] ; \quad \begin{bmatrix} u_{0,i}^{(y)} \\ u_{t,i}^{(y)} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma_{y,0}^2 & \sigma_{y,(0,t)} \\ \sigma_{y,(t,0)} & \sigma_{y,t}^2 \end{pmatrix} \right] ; \quad \begin{bmatrix} \boldsymbol{\epsilon}_i^{(x)} \\ \boldsymbol{\epsilon}_i^{(y)} \end{bmatrix} \sim \mathcal{N}_{2j}(\mathbf{0}, \mathbf{I}_{2j})$$

In order to implement this model within R-INLA a few tricks need to be used not previously discussed:

- INLA allows for random effects to be copied with a scaling factor γ into a different likelihood. For this one uses the following syntax:

  ```
  f(Intercept1)+ f(Intercept12, copy="Intercept1", hyper = list(beta = list(fixed=FALSE)))
  ```

  Hereby we indicate that we want to copy the element Intercept1 into a different likelihood with a non-fixed scaling factor *gamma*.

- To ensure that all elements being copied use the same scaling factor γ, the following syntax is used:

  ```
  f(x1, x)+
    f(x12, x, copy="x1", same.as = 'Intercept12', hyper = list(beta = list(fixed=FALSE)))
  ```

  Hereby we ensure that the element (x1) is copied with a scaling factor γ that is the same as was used when copying and scaling 'Intercept12'.

- A problem within INLA is that only random effects can be copied and scaled in this way. Thus, the only way to copy and scale fixed effects is to turn them into random effects with 2 levels, one for the endogenous covariate x and one for the outcome y. As example, we would have a random effect for the Intercept $\beta_{0,k} \sim \mathcal{N}(0, \sigma_0^2), \quad k = 1, 2$, with 2 levels, one for x and one for y, equal for all subjects. We are then not interested in the variance σ_0^2 of this random effect but instead in the realisation of the random effect for both the endogenous covariate y_1 and the outcome y_2.
  In this way all of the fixed effects are written down as random effects and copied with the same scaling parameter. Particularly within the Bayesian framework the difference between fixed and random effects is more subtle than in a frequentist approach, as both fixed and random effects have parameters that are random variables.

- Random effects are implemented as was discussed in definition 4.1, and copied in the same way as are fixed effects.

Definition 4.8 (Model 3A1) *Model 3A1 is a small modification from model 3A and the two models are nested. The only difference is that in model 3A1 there is no dependence between the random slope and random intercept in the endogenous covariate x. Thus, the covariance matrix of the random effects for the endogenous covariate x becomes:*

$$\begin{bmatrix} u_{0,i}^{(x)} \\ u_{t,i}^{(x)} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma_{x,0}^2 & 0 \\ 0 & \sigma_{x,t}^2 \end{pmatrix} \right]$$

This model was introduced to test, based on the goodness of fit measures, how similar these 2 nested models would perform on simulated data.

Definition 4.9 (Model 3B) *Model 3B was inspired by [Guo and Carlin, 2004], an article in which many models with scaled linear predictors are discussed. In Model 3B the random intercept and random slope are copied and scaled with different scaling parameters γ_1 and γ_2.*

$$\begin{cases} x_{i,j} = (\beta_0^{(x)} + u_{0,i}^{(x)}) + \beta_v^{(x)} \cdot v_i + \left(\beta_t^{(x)} + u_{t,i}^{(x)}\right) \cdot t_{i,j} + \epsilon_{i,j}^{(x)} \\ y_{i,j} = \gamma_1 \cdot u_{0,i}^{(x)} + \gamma_2 \cdot u_{t,i}^{(x)} + (\beta_0^{(y)} + u_{0,i}^{(y)}) + \beta_v^{(y)} \cdot v_i + \left(\beta_t^{(y)} + u_{t,i}^{(y)}\right) \cdot t_{i,j} + \epsilon_{i,j}^{(y)} \end{cases}$$

with

$$u_{0,i}^{(x)} \sim \mathcal{N}(0, \sigma_{x,0}^2); \quad u_{t,i}^{(x)} \sim \mathcal{N}(0, \sigma_{x,t}^2); \quad \begin{bmatrix} u_{0,i}^{(y)} \\ u_{t,i}^{(y)} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma_{y,0}^2 & \sigma_{y,(0,t)} \\ \sigma_{y,(t,0)} & \sigma_{y,t}^2 \end{pmatrix} \right]; \quad \begin{bmatrix} \boldsymbol{\epsilon}_i^{(x)} \\ \boldsymbol{\epsilon}_i^{(y)} \end{bmatrix} \sim \mathcal{N}_{2j}(\mathbf{0}, \mathbf{I}_{2j})$$

For the models with scaled linear predictor it is important to note that they can be rewritten into a form where there is no dependence between x and y. If we take Model 3A as example, we can write the linear predictor for y as:

$$\begin{aligned} y_{i,j} &= \gamma \cdot m_{ij} + (\beta_0^{(y)} + u_{0,i}^{(y)}) + \beta_v^{(y)} \cdot v_i + \left(\beta_t^{(y)} + u_{t,i}^{(y)}\right) \cdot t_{i,j} + \epsilon_{i,j}^{(y)} = \\ &= (\gamma\beta_0^{(x)} + \beta_0^{(y)}) + (\gamma\beta_v^{(x)} + \beta_v^{(y)})v_i + \left(\gamma\beta_t^{(x)} + \beta_t^{(y)}\right) \cdot t_{i,j} + (\gamma u_{0,i}^{(x)} + u_{0,i}^{(y)}) + (\gamma u_{t,i}^{(x)} + u_{t,i}^{(y)})t_{i,j} + \epsilon_{i,j}^{(y)} = \\ &= \beta_0^{(y)'} + \beta_v^{(y)'} v_i + \beta_t^{(y)'} t_{i,j} + u_{0,i}^{(y)'} + u_{t,i}^{(y)'} \cdot t_{i,j} + \epsilon_{i,j}^{(y)} \end{aligned}$$

with $\beta_0^{(y)'} = \gamma\beta_0^{(x)} + \beta_0^{(y)}$, $\beta_v^{(y)'} = \gamma\beta_v^{(x)} + \beta_v^{(y)}$ and $\beta_t^{(y)'} = \gamma\beta_t^{(x)} + \beta_t^{(y)}$.
For the random effects, they now have distribution:

$$\begin{bmatrix} u_{0,i}^{(y)'} \\ u_{t,i}^{(y)'} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \gamma^2\sigma_{x,0}^2 + \sigma_{y,0}^2 & \gamma^2\sigma_{x,(0,t)} + \sigma_{y,(0,t)} \\ \gamma^2\sigma_{x,(t,0)} + \sigma_{y,(t,0)} & \gamma^2\sigma_{x,t}^2 + \sigma_{y,t}^2 \end{pmatrix} \right]$$

Thus, although the models with the scaled linear predictor do seem to be very different from the independent joint models, the scaled linear predictor model can be rewritten in the same form as Model 0, see definition 4.1.
Further note that none of the models with scaled linear predictor can be fit using the packages nlme, lmer or MCMCglmm.

4.5 Summary of Models introduced

A summary of all the Models introduced in the previous sections is given in table 1. Note that all models have the same fixed part, consisting of:

$$\begin{cases} x_{i,j} = \beta_0^{(x)} + \beta_v^{(x)} \cdot v_i + \beta_t^{(x)} \cdot t_{i,j} + \epsilon_{i,j}^{(x)} \\ y_{i,j} = \beta_0^{(y)} + \beta_v^{(y)} \cdot v_i + \beta_t^{(y)} \cdot t_{i,j} + \epsilon_{i,j}^{(y)} \end{cases}$$

The subsequent random effects, errors and scaling elements unique to each model are given in table 1.

4.6 Joint Models and their implementation in different R packages

We have attempted to fit all models discussed so far using the R packages R-INLA, nlme, lmer and MCMCglmm. The packages R-INLA and MCMCglmm are Bayesian packages for Mixed Models, with R-INLA using the INLA framework for inference while MCMCglmm uses MCMC sampling. Both nlme and lmer are frequentist methods for solving Mixed Models. An overview of the different models and the packages that are able to fit them is given in table 2.

		R-INLA	NLME	LMER	MCMCglmm
Independent Mixed Models	Only Random Effects	✓	✓	✓	✓
Multivariate Joint Model	Correlated Residual Errors	✓	✓	✗	✓
Joint Mixed Models	Random effect + correlated residual errors	✓	✓	✗	✗
	Only random effects	✓	✓	✓	✓
Joint Mixed Models with scaled linear predictor	Only Random Effects	✓	✗	✗	✗

Table 2: Tabel indicating which R-packages can fit which joint models.

5 Simulation Study

In order to test the implementation of the different joint models in R we used simulated data. To construct the simulated data a total of $i = 1, ..., N$ patients were simulated with $j = 1, ..., n$ measurements per patient. Per measurement a variable v was randomly sampled from the $\mathcal{N}(0, 1)$ distribution, and the progression over time was modelled to be linear. In order to make the data unbalanced at every measurement time there was a probability p_1 that the endogenous covariate x was measured and a probability p_2 that the outcome y was measured. An example of such a data-set is shown in table 3.

Table 3: Table showing part of the simulated dataset

id	v	time	y_1 observed	y_2 observed
1	-0.67	0	1	0
1	-1.18	1	0	1
1	0.87	2	1	0
2	0.11	0	1	1
2	-1.36	1	0	1
2	-0.08	2	0	0

The endogenous covariate and the outcome were simulated according to the different models, with the addition of an extra error sampled from a $\mathcal{N}(0, 0.5)$ distribution.

5.1 Set-up of Simulation Study

Using the constructed dataset the models fitted in R-INLA were compared. We proceeded as follows:

- We simulated data as described in section 5. The endogenous covariate x and the outcome y were then simulated according to each of the models. As we are only interested in models with more than 2 measurements per subject, models 1A and 2A were not included. We thus simulated from all other models, equalling a total of 7 simulated data-sets.
- We then fitted each data-set with each of the models. We hereby calculated all goodness of fit measures available (see section 3.3), which include the Marginal Likelihood, DIC, WAIC, CPO, PIT and the three different types of MSE (see section 3.3.5). To test the correctness of the Marginal Likelihood and the DIC, we also approximated the DIC via $-2\log MLIK$, as this approximation is generally correct (ADD REFERENCE!!). This gave us an extra criteria to determine whether the goodness of fit measures supplied by INLA are correct.
- For each model we generated 5 datasets and averaged the goodness of fit measures over these 5 datasets. Hereby we thus performed a 5-fold Cross-validation. Since the data is generated we simply used this approach instead of generating a larger dataset and splitting it into 5 pieces.
- In total we thus fitted 7 (Datasets generated according to each Model) ×5 (Each dataset generated 5 times for 5-fold CV) ×7 (Each dataset generated is fit by each model) = 245 models.
- The parameters used are: $N = 75, n = 6, p_1 = 0.7, p_2 = 0.7, CV = 5, \boldsymbol{\beta} = (\beta_0^{(x)}, \beta_v^{(x)}, \beta_t^{(x)}, \beta_0^{(y)}, \beta_v^{(y)}, \beta_t^{(y)}) = (2.4, 2.5, 3, 1.5, 2.5)$.
  To calculate MSE_{same}, the MSE on subsequent measurements of subjects included in fitting the model (to determine how well the model fitted the random effects), an additional $n_2 = 4$ measurements per subject were sampled (but not used in the fitting process).
  To calculate the MSE on test subjects (to determine the marginal results) an additional $N = 25$ subjects with $n = 10$ measurements were sampled but not included in the fitting process.
- The exact parameters used to generate the endogenous covariate x and the outcome y according to each of the models are given in the appendix (TO DO, CREATE APPENDIX!!).

5.2 Results of Simulation Study

In table 4 we can see the results when data is simulated according to Model 0. In bold are shown the models performing best on the different goodness of fit metrics. We see that on most metrics the fit of Model 0 gives best results, which is to be expected as this is the model in which the data were simulated. We also notice that the majority of goodness of fit measures are very close together, indicating that the difference in goodness of fit between the models is not very large. One striking exception to this rule is Model 2A1, which is caused by the fact that Model 2A1 has no means of modelling a random slope and thus can not correctly fit the data at hand (which does include a random slope per subject).

Data simulated from Model 0

Fitted Model	Model_0	Model_2A1	Model_2B	Model_2B1	Model_3A	Model_3A1	Model_3B
MLIK	**-1169**	-1932	-1179	-1208	-1219	-1200	-1197
DIC_approx	**2339**	3864	2358	2415	2438	2400	2395
DIC	**1210**	3520	**1210**	1215	**1210**	**1210**	**1210**
WAIC	1196	3532	**1193**	1202	**1193**	1196	1194
PIT	0.0247	0.0880	0.0251	0.0249	**0.0245**	0.0247	0.0269
CPO	**677.9**	1774.3	680.8	679.7	679.3	678.1	679.9
MSE_train	0.331	10.789	0.334	0.350	0.330	0.337	**0.328**
MSE_same	1.57	98.08	1.59	1.69	1.56	1.61	**1.54**
MSE_others	127.77	128.03	**126.76**	127.77	128.25	128.23	127.77

Table 4: Results of simulation study when data is simulated according to Model 0. In bold are the models with best results according to the different goodness of fit metrics.

When examining table 5 we see that the fit according to Model 2A1 is best on almost all goodness of fit measures. An exception to this rule is the MSE on the training set, which can be explained by the fact that Models 0 and 3A1 can overfit the data by using a subject-specific random slope which in reality does not exist. The overfitting hypothesis seems to be confirmed by the fact that Model 2A1 performs best on MSE_{same}, a measure which takes into account the correctness of the inference of the random effects.

Data simulated from Model 2A1

Fitted Model	Model_0	Model_2A1	Model_2B	Model_2B1	Model_3A	Model_3A1	Model_3B
MLIK	-848	**-793**	-841	-847	-857	-877	-832
DIC_approx	1695	**1587**	1683	1695	1714	1755	1664
DIC	1113	**1073**	1112	1111	1084	1112	1083
WAIC	1104	**1079**	1105	1103	1082	1104	1082
PIT	0.027	0.028	0.027	0.030	**0.026**	0.027	0.026
CPO	588.1	**549.7**	586.6	584.8	565.0	587.5	565.0
MSE_train	**0.212**	0.232	0.213	0.213	0.221	**0.212**	0.221
MSE_same	0.53	**0.32**	0.52	0.51	0.44	0.54	0.44
MSE_others	2.67	2.67	2.67	2.67	2.67	2.67	2.67

Table 5: Results of simulation study when data is simulated according to Model 2A1. In bold are the models with best results according to the different goodness of fit metrics.

In table 6 the results when simulating data according to models 2B and 2B1 are shown. Over all the goodness of fit measures we see that the fitted model 2B performs best, even when data is generated according to model 2B1. We thus see that the nested model 2B1, where in comparison to model 2B some parameters are set to 0, does not give a better fit than model 2B. We would have expected for some goodness of fit parameters, such as the DIC and WAIC, which take into account the number of parameters used, to declare model 2B1 better. However, to our surprise even these measures do not indicate a better performance of model 2B1.

	Data simulated from Model 2B							**Data simulated from Model 2B1**						
Fitted Model	0	2A1	2B	2B1	3A	3A1	3B	0	2A1	2B	2B1	3A	3A1	3B
MLIK	-1178	-1924	**-1166**	-1204	-1223	-1208	-1201	-1167	-1919	**-1165**	-1203	-1218	-1196	-1199
DIC_approx	2356	3847	**2332**	2408	2446	2416	2403	2333	3838	**2331**	2406	2436	2393	2398
DIC	1216	3535	**1214**	1220	1214	1218	1214	1212	3499	**1209**	1293	1210	1214	**1209**
WAIC	1200	3545	1199	1205	**1198**	1202	1197	1197	3509	**1193**	1283	**1193**	1199	**1193**
PIT	0.026	0.084	0.026	**0.025**	0.027	**0.025**	0.026	**0.025**	0.076	0.026	0.030	0.028	**0.025**	0.028
CPO	685.9	1779.5	**683.1**	685.5	685.0	687.5	684.3	**679.0**	1762.5	679.3	709.6	680.2	680.1	679.7
MSE_train	**0.33**	11.15	**0.33**	0.34	**0.33**	0.34	**0.33**	0.36	10.70	**0.34**	0.43	0.35	0.36	0.35
MSE_same	1.55	100.41	**1.51**	1.60	1.54	1.59	1.53	1.71	95.98	**1.58**	2.01	1.62	1.69	1.66
MSE_others	118.80	119.15	118.84	118.86	118.75	**118.55**	118.79	135.48	135.89	135.49	135.49	135.43	**135.17**	135.46

Table 6: Results of simulation study when data is simulated according to Models 2B and 2B1. In bold are the models with best results according to the different goodness of fit metrics.

	Data simulated from Model 3A							Data simulated from Model 3A1						
Fitted Model	0	2A1	2B	2B1	3A	3A1	3B	0	2A1	2B	2B1	3A	3A1	3B
MLIK	-1228	-1989	-1184	-1278	-1264	-1205	-1228	-1205	-2012	-1182	-1224	-1252	-1190	-1223
DIC_approx	2456	3978	2367	2556	2528	2410	2455	2409	4023	2365	2448	2504	2381	2445
DIC	1217	3687	1205	1268	1213	1205	1208	1209	3700	1202	1302	1209	1200	1206
WAIC	1200	3698	1190	1262	1196	1190	1192	1194	3713	1186	1291	1192	1186	1190
PIT	0.024	0.091	0.027	0.025	0.027	0.027	0.027	0.024	0.075	0.027	0.025	0.025	0.025	0.027
CPO	688.4	1854.6	674.5	701.8	685.7	672.3	675.9	676.9	1862.8	672.3	709.3	681.1	666.3	675.7
MSE_train	0.58	17.13	0.33	0.55	0.65	0.33	0.56	0.49	16.63	0.33	0.38	0.59	0.33	0.54
MSE_same	3.15	154.05	1.52	2.84	3.63	1.50	2.96	2.58	148.68	1.51	1.76	3.20	1.50	2.83
MSE_others	171.29	173.62	171.18	171.17	171.82	171.47	171.30	232.78	232.94	232.76	236.13	233.24	233.16	232.82

Table 7: Results of simulation study when data is simulated according to Models 3A and 3A1. In bold are the models with best results according to the different goodness of fit metrics.

Data simulated from Model 3B

Fitted Model	Model_0	Model_2A1	Model_2B	Model_2B1	Model_3A	Model_3A1	Model_3B
MLIK	-1233	-2042	**-1183**	-1271	-1255	-1263	-1226
DIC_approx	2466	4083	**2366**	2541	2509	2526	2451
DIC	1217	3786	1210	1328	**1208**	1218	**1208**
WAIC	1201	3799	1195	1321	**1193**	1202	1194
PIT	**0.022**	0.085	0.026	0.024	0.025	0.024	0.026
CPO	690.2	1905.3	**677.7**	731.1	678.9	690.4	676.9
MSE_train	0.56	19.51	**0.35**	0.45	0.55	0.57	0.58
MSE_same	3.07	174.78	**1.65**	2.17	2.99	3.14	3.14
MSE_others	177.51	179.78	**176.92**	177.19	177.66	178.09	177.51

Table 8: Results of simulation study when data is simulated according to Models 3B. In bold are the models with best results according to the different goodness of fit metrics.

References

[Efron, 2011] Efron, B. (2011). Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing and Prediction. *International Statistical Review*, 79(1):126–127.

[Fitzmaurice,] Fitzmaurice, G. *Longitudinal Data Analysis*.

[Guo and Carlin, 2004] Guo, X. and Carlin, B. P. (2004). Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages. *American Statistician*, 58(1):16–24.

[P.J. Diggle, 2016] P.J. Diggle, P. H. (2016). Analysis of Longitudinal Data (Book). *Book*, (April):5–24.

[Rizopoulos, 2012] Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*, volume 9781447128.

[Rue et al., 2009] Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 71(2):319–392.

[Van Niekerk et al., 2019] Van Niekerk, J., Bakka, H., and Rue, H. (2019). Joint models as latent Gaussian models - not reinventing the wheel. (January).

[van Niekerk et al., 2021] van Niekerk, J., Bakka, H., and Rue, H. (2021). Competing risks joint models using R-INLA. *Statistical modelling*, 21(1-2):56–71.

[van Niekerk et al., 2019] van Niekerk, J., Bakka, H., Rue, H., and Schenk, O. (2019). New frontiers in Bayesian modeling using the INLA package in R. (2018):1–29.

Model	Random effects	Errors	Scaling
0	$\begin{bmatrix} u^{(x)}_{0,i} \\ u^{(x)}_{t,i} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma^2_{x,0} & \sigma_{x,(0,t)} \\ \sigma_{x,(t,0)} & \sigma^2_{x,t} \end{pmatrix}\right]$ $\begin{bmatrix} u^{(y)}_{0,i} \\ u^{(y)}_{t,i} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma^2_{y,0} & \sigma_{y,(0,t)} \\ \sigma_{y,(t,0)} & \sigma^2_{y,t} \end{pmatrix}\right]$	$\begin{bmatrix} \boldsymbol{\epsilon}^{(x)}_i \\ \boldsymbol{\epsilon}^{(y)}_i \end{bmatrix} \sim \mathcal{N}_{2j}(\mathbf{0}, \mathbf{I}_{2j})$	✗
1A	✗	$\begin{bmatrix} \epsilon^{(x)}_{i,1} \\ \epsilon^{(x)}_{i,2} \\ \epsilon^{(y)}_{i,1} \\ \epsilon^{(y)}_{i,2} \end{bmatrix} \sim \mathcal{N}_4 \left[\mathbf{0}, \boldsymbol{\Sigma}_i\right]$	✗
2A	$\begin{bmatrix} u^{(x)}_{0,i} \\ u^{(y)}_{0,i} \end{bmatrix} \sim \mathcal{N}_4 \left[\mathbf{0}, \begin{pmatrix} \sigma^2_{x,0} & \sigma_{(x,y),0} \\ \sigma_{(y,x),0} & \sigma^2_{y,0} \end{pmatrix}\right]$	$\begin{bmatrix} \epsilon^{(x)}_{i,1} \\ \epsilon^{(x)}_{i,2} \\ \epsilon^{(y)}_{i,1} \\ \epsilon^{(y)}_{i,2} \end{bmatrix} \sim \mathcal{N}_4 \left[\mathbf{0}, \boldsymbol{\Sigma}_i\right]$	✗
2A1	$\begin{bmatrix} u^{(x)}_{0,i} \\ u^{(y)}_{0,i} \end{bmatrix} \sim \mathcal{N}_4 \left[\mathbf{0}, \begin{pmatrix} \sigma^2_{x,0} & \sigma_{(x,y),0} \\ \sigma_{(y,x),0} & \sigma^2_{y,0} \end{pmatrix}\right]$	$\begin{bmatrix} \boldsymbol{\epsilon}^{(x)}_i \\ \boldsymbol{\epsilon}^{(y)}_i \end{bmatrix} \sim \mathcal{N}_{2j}(\mathbf{0}, \mathbf{I}_{2j})$	✗
2B	$\begin{bmatrix} u^{(x)}_{0,i} \\ u^{(y)}_{0,i} \\ u^{(x)}_{t,i} \\ u^{(y)}_{t,i} \end{bmatrix} \sim \mathcal{N}_4 \left[\mathbf{0}, \begin{pmatrix} \sigma^2_{x,0} & ... & ... & ... \\ ... & \sigma^2_{y,0} & ... & ... \\ ... & ... & \sigma^2_{x,t} & ... \\ ... & ... & ... & \sigma^2_{y,t} \end{pmatrix}\right]$	$\begin{bmatrix} \boldsymbol{\epsilon}^{(x)}_i \\ \boldsymbol{\epsilon}^{(y)}_i \end{bmatrix} \sim \mathcal{N}_{2j}(\mathbf{0}, \mathbf{I}_{2j})$	✗
2B1	$\begin{bmatrix} u^{(x)}_{0,i} \\ u^{(y)}_{0,i} \\ u^{(x)}_{t,i} \\ u^{(y)}_{t,i} \end{bmatrix} \sim \mathcal{N}_4 \left[\mathbf{0}, \begin{pmatrix} \sigma^2_{x,0} & ... & 0 & 0 \\ ... & \sigma^2_{y,0} & 0 & 0 \\ 0 & 0 & \sigma^2_{x,t} & ... \\ 0 & 0 & ... & \sigma^2_{y,t} \end{pmatrix}\right]$	$\begin{bmatrix} \boldsymbol{\epsilon}^{(x)}_i \\ \boldsymbol{\epsilon}^{(y)}_i \end{bmatrix} \sim \mathcal{N}_{2j}(\mathbf{0}, \mathbf{I}_{2j})$	✗
3A	$\begin{bmatrix} u^{(x)}_{0,i} \\ u^{(x)}_{t,i} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma^2_{x,0} & \sigma_{x,(0,t)} \\ \sigma_{x,(t,0)} & \sigma^2_{x,t} \end{pmatrix}\right]$ $\begin{bmatrix} u^{(y)}_{0,i} \\ u^{(y)}_{t,i} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma^2_{y,0} & \sigma_{y,(0,t)} \\ \sigma_{y,(t,0)} & \sigma^2_{y,t} \end{pmatrix}\right]$	$\begin{bmatrix} \boldsymbol{\epsilon}^{(x)}_i \\ \boldsymbol{\epsilon}^{(y)}_i \end{bmatrix} \sim \mathcal{N}_{2j}(\mathbf{0}, \mathbf{I}_{2j})$	$x_{i,j} = m_{ij} + \epsilon^{(x)}_{i,j}$ $y_{i,j} = \gamma m_{ij} + ...$
3A1	$\begin{bmatrix} u^{(x)}_{0,i} \\ u^{(x)}_{t,i} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma^2_{x,0} & 0 \\ 0 & \sigma^2_{x,t} \end{pmatrix}\right]$ $\begin{bmatrix} u^{(y)}_{0,i} \\ u^{(y)}_{t,i} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma^2_{y,0} & \sigma_{y,(0,t)} \\ \sigma_{y,(t,0)} & \sigma^2_{y,t} \end{pmatrix}\right]$	$\begin{bmatrix} \boldsymbol{\epsilon}^{(x)}_i \\ \boldsymbol{\epsilon}^{(y)}_i \end{bmatrix} \sim \mathcal{N}_{2j}(\mathbf{0}, \mathbf{I}_{2j})$	$x_{i,j} = m_{ij} + \epsilon^{(x)}_{i,j}$ $y_{i,j} = \gamma m_{ij} + ...$
3B	$\begin{bmatrix} u^{(x)}_{0,i} \\ u^{(x)}_{t,i} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma^2_{x,0} & 0 \\ 0 & \sigma^2_{x,t} \end{pmatrix}\right]$ $\begin{bmatrix} u^{(y)}_{0,i} \\ u^{(y)}_{t,i} \end{bmatrix} \sim \mathcal{N}_2 \left[\mathbf{0}, \begin{pmatrix} \sigma^2_{y,0} & \sigma_{y,(0,t)} \\ \sigma_{y,(t,0)} & \sigma^2_{y,t} \end{pmatrix}\right]$	$\begin{bmatrix} \boldsymbol{\epsilon}^{(x)}_i \\ \boldsymbol{\epsilon}^{(y)}_i \end{bmatrix} \sim \mathcal{N}_{2j}(\mathbf{0}, \mathbf{I}_{2j})$	$y_{i,j} = \gamma_1 u^{(x)}_{0,i} +$ $+\gamma_2 u^{(x)}_{t,i}$

Table 1: Summary of all models introduced in the previous sections. The residual errors covariance matrix $\boldsymbol{\Sigma}_i$ is the unstructured variance-covariance matrix.