Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

Aurora Christine Hofman

A shared Parameter Model Accounting for Dropout Not at Random

A Case Study on Blood Pressure in the HUNT Study

Master's thesis in applied mathematics and physics
Supervisor: Ingelin Steinsland
January 2022

NTNU
Norwegian University of
Science and Technology

Aurora Christine Hofman

A shared Parameter Model Accounting for Dropout Not at Random

A Case Study on Blood Pressure in the HUNT Study

Master's thesis in applied mathematics and physics
Supervisor: Ingelin Steinsland
January 2022

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

Abstract

In this work, we propose using a shared parameter model (SPM) in the Bayesian framework to account for missing data due to dropout in population-based health surveys. We use data from the longitudinal Trøndelag Health (HUNT) Study in a predictive model for systolic blood pressure with a modeling cohort consisting of 64385 participants out of which 43.1% dropped out. Further, an validation cohort is used to validate the model.

A novel evaluation scheme based on comparing the predictive performance of the fitted SPM with and without conditioning on the missing status is proposed. If the data are missing at random (MAR), there is no additional information in the missing process and no extra benefit of conditioning on the missing status.

The results demonstrate that the SPM is suitable for inference for a dataset of this size and structure and indicates that blood pressure missing due to dropout in the HUNT Study is missing not at random (MNAR). The SPM gives different parameter estimates than a naive model assuming data to be MAR. Through simulation studies based on models the suggested SPM and the naive model with MAR and MNAR data, we obtain indications that the SPM performs well on both MAR and MNAR data in contrast to the naive model, which only performs well when data is MAR. However, both models are biased when the data is MNAR.

SPM and naive models are compared based on predictive performance for the validation dataset. The naive model performs slightly better than the SPM when predicting blood pressure for the present participants. However, from the simulation study based on the SPM, we find that the naive model also performs better for the present participants, while the SPM performs better for the dropouts.

Through this work, we obtain strong indications that data MNAR should be accounted for when modeling systolic blood pressure using data from a longitudinal health survey. We also observe that blood pressure missing due to dropout in the HUNT Study is MNAR. Accounting for this gives a better representation of the data than assuming it to be MAR.

Sammendrag

I dette arbeidet foreslår vi å tilpasse en felles parametermodell (SPM) i et Bayesianske rammeverket for å ta hensyn til manglende data på grunn av frafall i befolkningsbaserte helseundersøkelser. Vi bruker data fra helse undersøkelse Trøndelag (HUNT) i en prediktiv modell for systolisk blodtrykk med en modelleringskohort bestående av 64385 deltagere hvorav 43.1% falt fra før påfølgende undersøkelse. Videre validerer vi modellene på et valideringsdatasett.

Vi foreslår en ny evaluaringsmetode basert på å sammenligne prediktiksjoner fra en tilpasset SPM med og uten å betinge på om deltagerne er tilstede eller ikke. Hvis dataen mangler tilfeldig (MAR) er det ingen ekstra informasjon i å vite manglende status og derfor heller ingen fordel å betinge på denne statusen.

Resultatene viser at en SPM er egnet for inferens på datasett av denne størrelse og struktur og indikerer at blood trykk manglende grunnet frafall i HUNT studien mangler ikke-tilfeldig (MNAR). SPM gir forskjellige parameterestimater sammenlignet med en naiv modell som antar at dataene er MAR. Gjennom simuleringsstudier der dataene er MNAR og MAR, får vi indikasjoner på at SPM presterer godt både når data er MNAR og MAR, i motsetning til den naive modellen, som kun presterer godt når data er MAR. Imidlertid gir begge modellene forventningsskjeve estimat når dataene er MNAR.

SPM og naive modeller blir sammenlignet basert på prediktive ferdigheter for valideringsdatasettet. Den naive modellen presterer litt bedre enn SPM når det gjelder å forutsi blodtrykket hos de tilstedeværende deltakerne. Vi finner imidlertid gjennom simuleringsstudier at SPM presterer bedre enn den naive modellen på deltakerne som ikke er tilstede når dataene er MNAR selv om den naive modellen presterer best på de tilstedeværende deltagerne.

Gjennom dette arbeidet får vi sterke indikasjoner på at det bør tas hensyn til data MNAR ved modellering av systolisk blodtrykk ved bruk av data fra longitudinelle helseundersøkelse. Vi observerer også at blood trykk manglende grunnet frafall i HUNT studien mangler ikke-tilfeldig. Å ta hensyn til dette gir en mer representativ modell enn vi får ved å anta at data mangler tilfeldig.

Preface

The work presented here was carried out at the Department of Mathematical Sciences in the autumn of 2021. It is a continuation of my project thesis conducted in the spring of 2021. Both these works are a continuation of the master thesis by Espeland [2020] who started researching how shared parameter models can be used to account for data missing not at random, on data from the Trøndelag Health (HUNT) Study. The HUNT Study is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology NTNU), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. Participation in the HUNT Study is voluntary, and all participants provided written informed consent before participation. The Regional Committee on Medical and Health Research Ethics of Norway (REK; 2018/1824) approved this work in July 2021.

This thesis has been directed by the wish to make the results based on Espeland and my research publicly available as a scientific journal paper. Therefore, the unique contributions of this thesis are directed by the research needed to finalize a scientific paper. Consequently, a substantial amount of time was used to reformulate and summarize previous findings and improve and generate new figures and tables. This process has been gratifying and has given me much insight into the world of academia. It let me learn a lot about the value of reproducing earlier results and iterating a work process to improve all aspects.

This work allowed me to write a conference paper based on the work done by Espeland [2020] together with Ingelin Steinsland and Emma Ingeström, giving me a crash course into the writing of scientific papers. This work was accepted, after being peer-reviewed, for a poster presentation at the 2021 Neural Information Processing Systems conference for the workshop themed "Your model is wrong" and is added to Appendix C.

Further, at the cusp of completing this thesis, Ingelin Steinsland, Lars Espeland, Emma Ingeström, and I are finalizing a larger scientific paper summarizing the work done by Espeland and myself. The self-contained paper draft is also added to Appendix C.

Writing two papers and a thesis has been very rewarding, although demanding at times. I am very proud of this being the final work of my master's degree at the Norwegian University of Science and Technology.

Trondheim, January 2022
Aurora Christine Hofman

Acknowledgement

The work presented here would not have been as good without the help of many.

First, I would like to thank my supervisor Ingelin Steinsland for excellent guidance, ideas, motivation, inspiration, and feedback. Thank you for helping me improve my academic skills, motivating me to contribute to the NORDSTAT conference in June 2021, and presenting a poster at the NeurIPS conference in December 2021. In addition, it has been very enlightening to collaborate on a larger scientific paper and obtain more insight into the world of academia.

I would also like to thank Emma Ingeström for providing the HUNT data, including writing the necessary applications allowing for this thesis and the corresponding paper, and ensuring the approval for presenting results at two conferences. Thank you for contributing with medical insight and having a great eye for orthography.

Also, thank you to the team at HUNT cloud.

Thank you, to all my great study colleagues, for providing great lunch conversations and making the time writing this thesis enjoyable.

Special thanks to my mom Kristin for spending endless hours correcting spelling mistakes and helping improve the flow of this work, in addition to providing constant support.

Thanks to Lucas, Sanne, and Tor-Ernst for being a wonderful family and always supporting me.

Thanks to all my great friends for making my time at the Norwegian University of Science and Technology truly amazing.

Finally, thank you, Sivert, just for being you!

Contents

1 Introduction

Data from population-based longitudinal health surveys are frequently used in medical research. In the common cohort study design where the same participants are followed over long periods, missing data due to loss to follow-up (dropout) in subsequent health surveys is almost inevitable. Proper handling of missing data is vital to obtain unbiased inference [Gad and Darwish, 2013, Little and Rubin, 2019, Chap. 1.3, 6].

This work aims to establish and validate a predictive model for systolic blood pressure using data from a longitudinal population-based health survey, the Trøndelag Health (HUNT) Study. Based on available literature [Anderson Jr et al., 1994, Whelton, 1994, Brown et al., 2000, Jiang et al., 2016, Espeland, 2020] we suggest a predictive model of future blood pressure (BP_F) based on initial blood pressure (BP_I), age (age), sex (sex), and body mass index (BMI). Elevated blood pressure is known to increase the risk of developing diseases related to the brain, heart, blood vessels, and kidney [Lewington et al., 2002, Tozawa et al., 2003, Rapsomaniki et al., 2014]. This medical condition affects more than 1.1 billion people annually. It accounts for over 10.8 million deaths per year, thereby surpassing smoking as the leading preventable cause of death for middle-aged and older adults worldwide [Zhou et al., 2017, Murray et al., 2020]. Early detection, prevention, and treatment of elevated blood pressure are of high priority in public health strategies [World Health Organizatoin, 2013]. Thus, obtaining unbiased, accurate models for BP_F is of great interest in medical research [Whelton, 1994].

The HUNT study consists of four consecutive surveys, HUNT1 (1984-86), HUNT2 (1995-97), HUNT3 (2006-08), and HUNT4 (2017-19), performed 11 years apart. We define a modeling cohort (HUNT2 cohort) consisting of 60385 participants with explanatory variables from HUNT2 and response values from HUNT3 and a validation cohort (HUNT3 cohort) consisting of 50201 participants with explanatory variables from HUNT3 and response values from HUNT4. The validation cohort is considered a new population with similar properties as the HUNT2 cohort. We chose this validation cohort to mimic the real scenario of using the models for predictive purposes. All participants have full records for the explanatory variables in the respective cohorts. As for most longitudinal population-based health surveys, a considerable proportion of the HUNT participants have dropped out, causing data to be missing at follow-up. In the HUNT2 cohort, 43.1% of the original participants have missing response values (lost to follow-up). In the HUNT3 cohort, 33.3% of the original participants are lost to follow-up.

Missing data can be categorized and described in terms of three missing processes; missing completely at random (MCAR), missing at random (MAR), and missing not at

random (MNAR)[Little and Rubin, 2019, Chap. 1.3, 6]. The two former processes are ignorable, meaning unbiased inference can be performed without modeling the missing process. However, the latter process is non-ignorable, and the missing process must be modeled simultaneously with the original model to obtain unbiased inference [Little and Rubin, 2019, Chap. 1.3, 6]. The three processes differ in how the observed and missing data affect the probability of having missing data. The data is MCAR if the probability of missing a data point is independent of the observed and the unobserved data. It is unreasonable that the probability of dropping out does not depend on age. Hence, data MCAR is not considered. The data is MAR if the probability of missing is conditional on the observed data and independent of the unobserved data. A model for the missing process with the same explanatory variables as the blood pressure is an example of a model assuming data MAR. Data that is neither MCAR nor MAR is MNAR [Gad and Darwish, 2013, Little and Rubin, 2019, Chap. 6]. If the part of the BP_F which can not be explained by age, sex, BP_I, and BMI affects the probability of dropping out, the data is MNAR. This part can be thought of as (unknown) explanatory variables not included in the models. It is reasonable that other health-related variables influence both the blood pressure and the probability of dropping out. Thus, we argue that a predictive model for BP_F should consider that data might be MNAR.

Ignoring missing data permits the use of powerful tools such as Maximum Likelihood [Dempster et al., 1977] and Multiple Imputation [Rubin, 1976]. However, missing data may contain important information about the true data distribution and should be treated with caution. [Little and Rubin, 2019, Chap. 1.3]. Missing data should only be ignored in cases where the data is MCAR or MAR [Little and Rubin, 2019, Mohan and Pearl, 2021].Ignoring data MNAR is known to cause biased inference [Little and Rubin, 2019, Chap. 1.3, 6]. Even though the assumption of data MAR is often not fulfilled, many of the available software packages and methods described in the literature assume data to be MAR [Balakrishnan, 2009, Rhoads, 2012, Little and Rubin, 2019, Mohan and Pearl, 2021]. However, several studies, especially in biostatistics, have accounted for missing data under the assumption of data MNAR. [Wu and Carroll, 1988, Little, 1993, Diggle and Kenward, 1994, Follmann and Wu, 1995, Little, 1995, Albert and Follmann, 2000, Molenberghs et al., 2008, Howe et al., 2016]. Popular choices for models accounting for data MNAR include the pattern mixture model, selection model, and shared parameter model (SPM) [Heckman, 1979, Wu and Carroll, 1988, Little, 1993, Henderson et al., 2000, Linero and Daniels, 2018, Little and Rubin, 2019, Chap. 15.4]. The pattern mixture model can be interpreted as a mixture of models for different populations characterized by their missingness pattern. The selection model is based on a factorization modeling the missing process conditional on the response together with the marginal distribution of the response [Gad and Darwish, 2013]. The SPM is based on the idea of a commonly shared variable affecting both the measurement process and the missing process. Given this variable, the two marginal densities are conditionally

independent. The SPM has been used to model longitudinal data subject to MNAR in several studies [Wu and Carroll, 1988, Follmann and Wu, 1995, Thomas et al., 1998, Pulkstenis et al., 1998, Vonesh et al., 2006, Creemers et al., 2010].

In this work, we propose a Bayesian SPM for future blood pressure. The model fits the framework of Bayesian latent Gaussian models and is suitable for Bayesian inference using Integrated Nested Laplace Approximations (INLA) [Rue et al., 2009, 2017, Gómez-Rubio, 2020]. Through INLA, computationally efficient inference is available. INLA uses numerical approximations to obtain the posterior marginal distribution of all parameters and hyperparameters. For a thorough description, the reader is encouraged to read the original and follow-up paper by Rue et al. [2009, 2017] and the paper by Martino and Riebler [2019]. Further, we compare the SPM with a naive model assuming the data to be MAR.

Molenberghs et al. stated that "each MNAR model fit to a set of observed data can be reproduced exactly by a MAR counterpart" [Molenberghs et al., 2008, p. 371]. Hence, the choice between models eventually comes down to choosing the most likely model assumptions [Enders, 2011]. Recent research has proven that taking the approach of causal modeling of graphical models by formulating the models through missingness graphs can give understanding and performance guarantees [Mohan and Pearl, 2021].

To the best of our knowledge, the literature provides little insight into practical validation of model performance on data MNAR. The current standard seems to be the use of simulation studies to check the reproducibility of the model. i.e., how well the original parameters are reproduced on simulated data, and sensitivity analysis to check the robustness of the models [Enders, 2011, Steinsland et al., 2014, Kaciroti and Little, 2021]. A case study is often presented fitting one or several models [Enders, 2011, Steinsland et al., 2014].

In this work, we validate our methods on a validation dataset with similar properties as the modeling data following two approaches. First, we compare the predictive performance of the SPM with the performance of a naive model, assuming the data to be MAR. We find the predictive distributions of all participants and compare the two models through the proper scoring rules [Gneiting and Raftery, 2007] continuous ranked probability score (CRPS) and Brier score. Second, we propose a new method to evaluate if the data is MNAR when following the SPM by comparing model predictions conditioned on missing status with model predictions without knowledge about the missing status. The key idea is that if data are MNAR, the missing status has information about the quantity of interest. Therefore we compare the predictive performance of the SPM with and without conditioning on missing status.

Reproducibility is of vital importance to drive science forwards. Since the data used in this work contains personal information, the original data can not be publicly available.

Therefore we have simulated a dataset that attempts to capture some of the same structures as the original dataset to make the reproducibility process as straightforward as possible. In addition, all code used in this work and a thorough description on how to use it is publicly available through GitHub [Hofman, 2021a]. In addition to providing the means to reproduce results on identical data, having open-source code simplifies the process of replication on new data as well. The paper by Ioannidis [2005] started a replication crisis within the psychology field. This paper concluded that the bias toward publishing positive findings rather than negative ones severely impacted the percentage of false significant findings. It also pointed out how academia stimulates new research much higher than replicating old findings.

This work proposes a SPM framework accounting for data MNAR due to dropout in cohorts studies. First, we propose a Bayesian shared parameter model for BP_F based on a longitudinal population-based health survey (the HUNT Study). We demonstrate that inference is computationally feasible. Second, we explore the difference between accounting for the possibility that data may be MNAR and assuming data MAR by comparing the SPM with a naive model assuming the data to be MAR. Third, we propose two strategies for validating the proposed models by exploring the predictive performance of the models and examining the likeliness of the data being MNAR for a realistic validation dataset. We perform a prior sensitivity study and simulation studies under different assumptions to further understand the proposed model and evaluation schemes for the given study system. The methodology provided applies to many other health indicators and other similar situations.

We build this work on the thesis by Espeland [2020] and my project thesis [Hofman, 2021b]. Espeland started researching the possibilities of using a SPM to account for data MNAR on data from HUNT1 and HUNT2 (HUNT1 cohort). He studied a broader specter of variables and therefore cleaned the data with respect to all these variables. Not all of the variables were significant. However, the dataset was not cleaned again. Hence the dataset used was smaller than strictly necessary. Espeland fitted a SPM, and a naive model for BP_F with age, BP_I, and sex as explanatory variables. In addition, he did a small simulation study and a small prior sensitivity analysis. He also briefly explored the possibility of adding BMI as an explanatory variable. My project thesis [Hofman, 2021b] extended the work by Espeland [2020] by cleaning the data only with respect to the variables used in the analysis. Hence, we use a larger dataset. In addition, all the data was standardized. The INLA software used for the analysis works better on standardized data. We noticed this in the form of fewer computational issues. Further, the model made on the HUNT1 cohort was used to predict the BP_F for the HUNT2 cohort, considered an independent dataset. These predictions were used to validate the model. The naive model in this work only modeled the BP_F. Hence, the predictions of missingness were compared to the average percentage of participants dropping out

between HUNT1 and HUNT2. This assumes the participants are missing completely at random, which we know not to be the case.

This work extends previous work by using the most recent HUNT data. I.e., the models are fitted to the HUNT2 cohort and used to predict BP_F in the HUNT3 cohort. We also naively model the dropout process. This addition enables comparisons of the parameter estimates and the posterior predictive distributions for the dropout process. Further, we propose a novel method to validate the MNAR assumption, for data following the SPM, by predicting BP_F at HUNT4 conditioning on the missing status at HUNT4. A prior sensitivity study is also performed, researching more possibilities than the prior sensitivity analysis done by Espeland [2020]. In addition, extensive simulation studies exploring the bias and coverage of the models are performed in addition to simulation studies on the two proposed validation schemes. Finally, we have created an open-source GitHub repository with all the code, a simulated dataset, and a detailed description of how to run the code to enable reproducibility of our work and comply with today's research standards.

This work is structured as follows: Chapter 2 provides background information regarding Bayesian Inference, Latent Gaussian Models, Gaussian Markov random fields, integrated nested Laplace approximations (INLA), missing data mechanisms and models commonly used, and briefly scoring rules. This chapter is inspired by the background theory section of my project thesis [Hofman, 2021b]. Section 3.1 introduces the blood pressure case study. First, we describe and explore the data used from the HUNT Study. Second, we set up the model specifications for the SPM and naive model with methods for inference and validation. The results from the case study are presented in Chapter 4. Chapter 5 consists of several simulation studies. We perform four simulation studies on data MNAR and MAR using models following both the SPM and naive model. In addition, we perform simulation studies targeting the proposed validation schemes. Chapter 6 contains a prior sensitivity analysis. To make this work accessible and reproducible, we provide all the code used in this work. In addition, we present both individual and population-based validation results. Chapter 7 accounts for where to find this code. It also gives a thorough account of how we create a simulated dataset where the explanatory and response variables are simulated. Chapter 8 summarizes the findings and discusses some of the challenges encountered in this work.

Background Theory

2

The models used in this work are all fitted in the Bayesian framework using integrated nested Laplace approximations (INLA) [Rue et al., 2009]. INLA can only be used on models belonging to the class of latent Gaussian models where the latent field is a Gaussian Markov random field. This chapter introduces the concepts of Bayesian inference. Further, we introduce latent Gaussian models, Gaussian Markov Random Fields, INLA, and present some missing data mechanisms and commonly used models to handle missing data. Finally, we introduce the continuous ranked probability score (CRPS) and the Brier score used for validation.

2.1 Bayesian Inference

There are two main frameworks within the field of statistics, the frequentist and the Bayesian. The frequentist framework is the most widespread and usually the first taught in statistics courses. The two frameworks differ in how they view the unknown model parameters. The frequentists believe that there exists one true value for every parameter in a model, and the objective is to find this one value. The Bayesianists believe the underlying model parameters have a distribution, not a single true value. The objective then becomes to find this underlying distribution of the model parameters. This framework allows expert knowledge to be incorporated into the model through a prior distribution for model parameters. The process of modeling within the Bayesian framework is as follows. First, we assign a prior distribution $\pi(\cdot)$ to all model parameters. This distribution can be informative/narrow, i.e., allowing for expert or prior knowledge to be incorporated into the model, or the distribution can be uninformative/wide, i.e., we have no prior knowledge of the model parameters. After assigning a prior distribution, we use the observed data $\boldsymbol{y}$ to obtain the likelihood function $\pi(\boldsymbol{y}|\cdot)$ for the model parameters given the observed data. Finally, we combine this information to obtain the posterior distribution of the model parameters, $\pi(\cdot|\boldsymbol{y})$ using Bayes rule:

$$\pi(\cdot|\boldsymbol{y}) = \pi(\boldsymbol{y}|\cdot)\pi(\cdot).$$

We illustrate the difference through an example. A coin manufacturer produces one million coins, and we want to find the probability of obtaining heads after flipping one of these coins. We assume the probability of obtaining heads is Bernoulli distributed with parameter p. A frequentist would then flip n of these coins and record the result of the flip for each coin. We can then estimate p by the proportion of coins resulting in heads.

For a frequentist, it would not matter if this experiment had been performed earlier by a different person. On the other hand, a Bayesian statistician would be very interested in the previous results and incorporate these results into the prior distribution of p, $\pi(p)$. Let us assume the previous experiment showed the coins have a slight bias towards heads. The Bayesianist could incorporate this by assigning a slightly asymmetrical distribution to this prior. Then the experiment is performed the same way as in the frequentist case. This observed data is incorporated into the likelihood and combined with the prior distribution results in a posterior distribution for p. I.e., p is no longer a single value but a distribution. This example does not contain any hyperparameters, and the latent field consists of the parameter p. More complex models such as hierarchical models will commonly need hyperparameters as well. These are parameters defining the distribution of the latent field. For example, instead of having one factory producing coins, we have $n = 1 : 10$ factories. Then we could be interested in finding p_n for the specific factory. Assuming p_n is beta distributed with parameters α and β, these would be the hyperparameters shared between all factories.

Occasionally it is feasible to obtain the posterior distribution analytically. However, in most cases, it is analytically intractable. Traditionally simulations are used to avoid this problem through Markov chain Monte Carlo simulations [Metropolis et al., 1953, Hastings, 1970]. The broad idea is to sample from the posterior distribution, usually indirectly. It can be proven that as long as the distribution is irreducible and aperiodic, the MCMC simulations always converge to the true sample given sufficient time Brooks [1998]. The most significant benefit of MCMC is its flexibility to fit almost any model. With the help of modern technology and computational resources available, it is a very powerful tool. The disadvantages are, among others, the lack of guarantee of convergence of a MCMC chain. Hence, it can be very time-consuming and computationally intensive for complex systems to obtain the posterior distribution with high certainty. In addition, it can be cumbersome to formulate a good MCMC algorithm. However, powerful tools such as STAN and WinBUGS make this easier [Spiegelhalter et al., 2003, Carpenter et al., 2017].

An alternative to the traditional MCMC is Integrated nested Laplace approximations (INLA) [Rue et al., 2009, Martins et al., 2013]. INLA is a fast and accurate methodology for approximating Bayesian inference without sampling. INLA and its limitations are described in more detail in Section 2.4

2.2 Latent Gaussian Models

The reader is assumed to be familiar with general linear models, general additive models, and general linear (additive) mixed models. These are all covered in Fahrmeir et al. [2007].

Latent Gaussian models fall within a subclass of the structured additive regression models [Rue et al., 2009]. This means the response y_i belongs to the class of exponential families. Within the class of exponential family the mean $E(y_i) = \mu$ is linked to a structured additive predictor η through a link function $h(\mu)$ such that $h(\mu) = \eta$. For the structured additive regression models η is defined as follows [Fahrmeir et al., 2007],

$$h(\mu) = \eta = \alpha + \sum_{k=1}^{n_\beta} \beta_k z_k + \sum_{j=1}^{n_f} f^{(j)}(u_j) + \epsilon.$$

Here $\{\beta_k\}$ represents the linear effects of covariates $\boldsymbol{z}$, $\{f^{(j)}(.)\}$ represents unknown functions of covariates $\boldsymbol{u}$, and ϵ is an unstructured term. To belong to the class of latent Gaussian model the prior distributions of α, $\{\beta_k\}$, $\{f^{(j)}(.)\}$ and ϵ must be Gaussian.

2.3 Gaussian Markov Random Fields

A Gaussian Markov random field (GMRF) is a random vector following a multivariate normal distribution in addition to some conditional independent assumptions [Rue and Held, 2005]. We follow the notation by Rue and Held [2005] for the formal definition. Let $\boldsymbol{x} = (x_1, ..., x_n)^T$ be normally distributed with mean $\boldsymbol{\mu}$ and covariance matrix Σ. Define the labeled grapf $G = (V, E)$ such that $V = (1, ..., n)$ and E consists of all pairs of nodes who are dependent on each other. I.e there is no edge between node i and j if x_i is independent of x_j given all other nodes, $x_i \perp x_j | \boldsymbol{x}_{-\{i,j\}}$.

The multivariate normal distribution has the following density function which can be rewritten to include the precision matrix Q instead of Σ.

$$\begin{aligned} f(\boldsymbol{x}) &= (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} exp\left[-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right] \\ &\Longleftrightarrow \\ f(\boldsymbol{x}) &= (2\pi)^{-\frac{n}{2}} |Q|^{\frac{1}{2}} exp\left[-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^T Q (\boldsymbol{x} - \boldsymbol{\mu}) \right]. \end{aligned} \tag{2.1}$$

We note that for $\boldsymbol{x}$ normally distributed with mean $\boldsymbol{\mu}$ and precision matrix Q then for every $i \neq j | \boldsymbol{x}_{-\{i,j\}}$ if and only if $Q_{ij} = 0$ [Rue and Held, 2005]. In other words we can see from Q weather x_i and x_j are conditionally independent.

Now we can state the formal definition of a GMRF.

Definition: A random vector $\boldsymbol{x} = (x_1, ..., x_n)^T \in \mathbb{R}^n$ is a GMRF with respect to a labelled graph $G = (V, E)$ with mean $\boldsymbol{\mu}$ and precision matrix $Q > 0$ if and only if its density has the following form

$$f(\boldsymbol{x}) = (2\pi)^{-\frac{n}{2}} |Q|^{\frac{1}{2}} exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T Q(\boldsymbol{x} - \boldsymbol{\mu}) \right]$$

and

$$Q_{ij} \neq 0 \quad \iff \{i, j\} \in E \quad \forall i \neq j$$

If most of the components of $\boldsymbol{x}$ are conditionally independent, then $\boldsymbol{x}$ is a GMRF with a sparse precision matrix, meaning most of the elements in the matrix are zero. This can be of great computational benefit because we do not have to store the zero entries in the precision matrix, making the computational time of computations involving the precision matrix much lower [Rue and Held, 2005].

2.4 Integrated Nested Laplace Approximation

Integrated nested Laplace approximation (INLA) is a state-of-the-art alternative to traditional Markov chain Monte Carlo simulations. It offers efficient numerical approximations for the marginal posterior distribution of all model parameters [Rue et al., 2009, Martino and Riebler, 2019] and exploits accurate numerical approximations and efficient numerical integration. INLA is restricted to the class of latent Gaussian models described in Section 2.2 where in addition, the latent field is a GMRF with a sparse precision matrix, introduced in Section 2.3, i.e., endowed with some Markovian properties.

Before we briefly introduce how INLA works we start by introducing Laplace approximations, which builds on the Taylor expansion given as follows,

$$f(x) = f(a)(x - a) + f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2 + ...$$

for an expansion of x around a point a. The Laplace approximation, $\tilde{\pi}_G(x)$, can be used to approximate the logarithm of any Gaussian distribution $\pi(x)$ by the three first terms of the Taylor expansion around the distribution mode $\hat{x}$ as follows,

$$\log(\pi(x)) \approx \log(\pi(\hat{x})) + \frac{\partial \log(\pi(\hat{x}))}{\partial x}(x - \hat{x}) + \frac{\partial^2 \log(\pi(\hat{x}))}{\partial x^2}(x - \hat{x})^2 \tag{2.2}$$

We know that $\frac{\delta \log(\pi(x))}{\delta x}\Big|_{x=\hat{x}} = 0$ by definition, hence (2.2) is equivalent to

$$\log(\pi(x)) \approx \log(\pi(\hat{x})) + \frac{\partial^2 \log(\pi(\hat{x}))}{\partial x^2}(x-\hat{x})^2 \tag{2.3}$$

If we define $\hat{\sigma} = -1/\frac{\partial^2 \log(\pi(\hat{x}))}{\partial x^2}$, then 2.3 can be rewritten to

$$\log(\pi(x)) \approx \log(\pi(\hat{x})) - \frac{1}{2\hat{\sigma}^2}(x-\hat{x})^2 \tag{2.4}$$

which is equivalent to

$$\pi(x) \approx \pi(\hat{x}) \exp\left[-\frac{1}{2\hat{\sigma}^2}(x-\hat{x})^2 \right] = \tilde{\pi}_G(x). \tag{2.5}$$

We note that $\exp\left[-\frac{1}{2\hat{\sigma}^2}(x-\hat{x})^2 \right]$ is the kernel of a normal distribution with mean $\hat{x}$ and standard deviation $\hat{\sigma}$. Hence,

$$\int_a^b \pi(x)dx \approx \pi(\hat{x})\sqrt{2\pi\hat{\sigma}^2}(\phi(b) - \phi(a)) \tag{2.6}$$

where $\phi(.)$ is the cumulative density function of the normal distribution with mean $\hat{x}$ and standard deviation $\hat{\sigma}$. Thus, we have a mean to approximate the integral of a normal distribution.

Let $\boldsymbol{x} = (x_1, .., x_n)$ be the set of all latent variables, all having Gaussian priors in addition to being a GRMF with a sparse precision matrix. In the notation used for latent Gaussian models, Section 2.2, $\boldsymbol{x} = (\alpha, \{\beta_k\}, \{f^{(j)}(.)\}, \epsilon_i)$. Let $\boldsymbol{y}$ be the observed data and $\boldsymbol{\theta}$ the set of hyperparameters.

The objective is to obtain the marginal posterior distribution $\pi(x_i|\boldsymbol{y})$ for all $i = 1, ..., n$ variables in the latent field. This is equivalent to the following

$$\pi(x_i|\boldsymbol{y}) = \int \pi(x_i, \boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta} = \int \pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})\pi(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}. \tag{2.7}$$

I.e we need $\pi(\boldsymbol{\theta}|y)$ and $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$.

$\pi(\boldsymbol{\theta}|y)$ can be computed by using the Laplace approximation and Bayes rule as follows,

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})} \approx \frac{\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}\Bigg|_{\boldsymbol{x}=\hat{\boldsymbol{x}}(\boldsymbol{\theta})} = \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}). \tag{2.8}$$

where $\hat{\boldsymbol{x}}(\boldsymbol{\theta})$ is the mode of $\boldsymbol{x}$.

We find $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ through an approximation. The approximation can be obtained in three different ways [Rue et al., 2009, Martino and Riebler, 2019]. The fastest is through the reuse of the Gaussian approximation $\pi_G(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ already computed to solve (2.8). This method is the least accurate of the three. The second option is to perform Laplace approximations a second time. This option is computationally demanding and reduces the speed. However, the results are very accurate. The third option is to perform a simplified Laplace approximation correcting $\pi_G(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ for local skewness. This option is less accurate than option two, although often sufficiently accurate and very computationally efficient. For further elaboration, the reader is encouraged to read the original paper by Rue et al. [2009].

The final INLA algorithm is given as follows [Martino and Riebler, 2019],

- Select points $k = 1, \ldots, K$ in the space of $\boldsymbol{\theta}$ where the density of $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ is high.

- Compute $\pi(\theta_k|\boldsymbol{y})$ using (2.8) for all k.

- For all θ_k compute $\pi(x_i|\theta_k, \boldsymbol{y})$ using one of the three aforementioned alternatives.

- Approximate (2.7) by

$$\tilde{\pi}(x_i|\boldsymbol{y}) = \sum_k \tilde{\pi}(x_i|\theta_k, \boldsymbol{y})\tilde{\pi}(\theta_k|\boldsymbol{y})\Delta_k. \tag{2.9}$$

Δ_k are appropriate weigths depending on how the K points where chosen. We note that the number of parameters must be small $n < 20$ for this scheme to be numerically solvable. For a more detailed explanation of the INLA algorithm, see Martino and Riebler [2019].

2.5 Missing Data

Missing data is a well-known challenge in most statistical research. As aforementioned, depending on the process behind the missing values, this has to be accounted for differently. This work divides missing processes into three categories, missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) as done in [Little and Rubin, 2019, Chap 1.5, 6]. We follow the notation given by [Little and Rubin, 2019, Chap 6]

Let $\boldsymbol{y_i}$ be the set of j measurements on the ith subject. Then $\boldsymbol{y_i}$ can be divided into an observed part $\boldsymbol{y_{io}}$ and a missing part $\boldsymbol{y_{im}}$, $\boldsymbol{y_i} = (\boldsymbol{y_{io}}, \boldsymbol{y_{im}})$. Let $\boldsymbol{m_i}$ be the vector of,

$$m_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is missing} \\ 0 & \text{otherwise.} \end{cases}$$

Then the full conditional of $\boldsymbol{y_i}$ and $\boldsymbol{m_i}$ is given as follows,

$$g(\boldsymbol{y_{io}}, \boldsymbol{y_{im}}, \boldsymbol{m_i}|\boldsymbol{\theta}, \boldsymbol{\psi}) \tag{2.10}$$

where the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ describes the measurement process and missing process, respectively [Gad and Darwish, 2013, Little and Rubin, 2019, Chap. 6.2]. This allow us to formally define data MCAR, MAR, and MNAR. The data is MCAR if the missing process is not affected by any observed or unobserved data i.e,

$$g(\boldsymbol{m_i}|\boldsymbol{y_{io}}, \boldsymbol{y_{im}}, \boldsymbol{\psi}) = g(\boldsymbol{m_i}|\boldsymbol{\psi}). \tag{2.11}$$

The data is MAR if the missing process can be influenced by the observed data but not the unobserved data;

$$g(\boldsymbol{m}|\boldsymbol{y_{io}}, \boldsymbol{y_{im}}, \boldsymbol{\psi}) = g(\boldsymbol{m_i}|\boldsymbol{y_{io}}, \boldsymbol{\psi}). \tag{2.12}$$

If the data is neither MCAR nor MAR, the data is, by definition, MNAR. The MCAR and MAR processes are ignorable, and it is possible to perform valid inference by modeling the measurement process alone [Little and Rubin, 2019, Chap. 6.2]. In contrast, if data is MNAR, the missing process must be modeled simultaneously with the measurement process to obtain unbiased inference [Little and Rubin, 2019, Chap. 6.2].

We illustrate the differences between data MCAR, MAR, and MNAR through a small simplified example. We want to model BP_F is with BP_I, age, sex, and BMI as the explanatory variable. Hence $y_i = BP_{Fi}$ and $m_i = 1$ if BP_{Fi} is missing, otherwise $m_i = 0$. If $p(m_i|age_i, sex_i, BP_{Ii}, BMI_i, BP_{Fi}) = p(m_i)$ the data is MCAR. I.e., the probability of missing BP_{F_i} is completely independent of the explanatory variables and the response. If $p(m_i|age_i, sex_i, BP_{Ii}, BMI_i, BP_{Fi}) = p(m_i|age_i, sex_i, BP_{Ii}, BMI_i)$ the data is MAR. When the data is MAR, the explanatory variables can affect the probability of dropout, i.e., when the elderly drop out because they are old. This is a slight simplification as the observed values of BP_F can also affect the missing process in the MAR case. When BP_{Fi} itself affects the probability of dropout, the data is MNAR. For example, if a participant suddenly increases blood pressure levels severely leading to a very poor health, enabling further attendance in surveys. This can be thought of as (unknown) explanatory variables not included in the models affecting both BP_F and probability of drop out. These three cases are illustrated in Figure 2.1.

We can see from (2.11) and (2.12) that MCAR is a special case of MAR. We will now briefly show why MAR is an ignorable process following the notation by [Little and Rubin, 2019, Chap. 6]. First we introduce the full likelihood, L_{full}, and the ignorable likelihood L_{ign} based on the observed values $(\boldsymbol{y_o}, \boldsymbol{m})$.

$$L_{full}(\boldsymbol{\theta}, \boldsymbol{\psi}|\boldsymbol{y_o}, \boldsymbol{m}) = \int f_Y(\boldsymbol{y_o}, \boldsymbol{y_m}|\boldsymbol{\theta}) f_{M|Y}(\boldsymbol{m}|\boldsymbol{y_o}, \boldsymbol{y_m}, \boldsymbol{\psi}) d_{\boldsymbol{y_m}}. \tag{2.13}$$

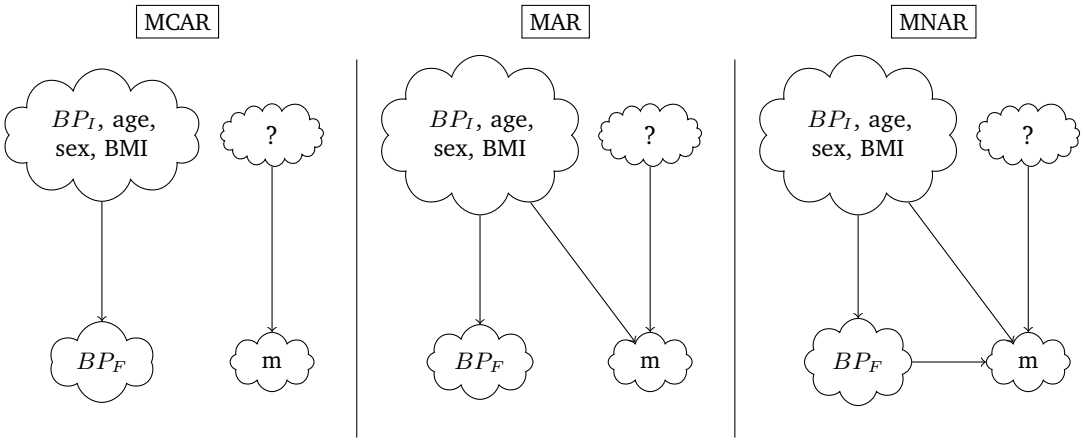


Figure 2.1.: Illustration of data missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) when modeling systolic blood pressure BP_F with BP_I, age, sex, and BMI as explanatory variables. (Modified version of Figure 1 in Hofman [2021b])

$$L_{ign}(\boldsymbol{\theta}|\boldsymbol{y_o}) = \int f_Y(\boldsymbol{y_o}, \boldsymbol{y_m}|\boldsymbol{\theta}) d_{\boldsymbol{y_m}}. \tag{2.14}$$

The formal definition of an ignorable missing process is given as follows [Little and Rubin, 2019]:

Definition: A missing process is ignorable if the following two conditions are fulfilled:

- The parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are distinct. I.e., the joint parameter space for $(\boldsymbol{\theta}, \boldsymbol{\psi})$ is the product of the individual parameter spaces for $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$.

- The full likelihood factors as

$$L_{full}(\boldsymbol{\theta}, \boldsymbol{\psi}|\boldsymbol{y_o}, \boldsymbol{m}) = L_{ign}(\boldsymbol{\theta}|\boldsymbol{y_o}) L_{rest}(\boldsymbol{\psi}|\boldsymbol{y_o}, \boldsymbol{m}) \tag{2.15}$$

for all $(\boldsymbol{\theta}, \boldsymbol{\psi})$ in the joint parameter space of $(\boldsymbol{\theta}, \boldsymbol{\psi})$.

For a MAR process we have independent $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ and

$$\begin{aligned} f(\boldsymbol{y_o}, \boldsymbol{m}|\boldsymbol{\theta}, \boldsymbol{\psi}) &= \int f(\boldsymbol{y_o}, \boldsymbol{y_m}, \boldsymbol{m}|\boldsymbol{\theta}, \boldsymbol{\psi}) d_{\boldsymbol{y_m}} \\ &= \int f_{M|Y}(\boldsymbol{m}|\boldsymbol{y_o}, \boldsymbol{y_m}, \boldsymbol{\theta}, \boldsymbol{\psi}) f(\boldsymbol{y_o}, \boldsymbol{y_m}|\boldsymbol{\theta}, \boldsymbol{\psi}) d_{\boldsymbol{y_m}} \\ &= \int f_{M|Y}(\boldsymbol{m}|\boldsymbol{y_o}, \boldsymbol{\psi}) f(\boldsymbol{y_o}, \boldsymbol{y_m}|\boldsymbol{\theta}, \boldsymbol{\psi}) d_{\boldsymbol{y_m}} \\ &= f_{M|Y}(\boldsymbol{m}|\boldsymbol{y_o}, \boldsymbol{\psi}) \int f(\boldsymbol{y_o}, \boldsymbol{y_m}|\boldsymbol{\theta}) d_{\boldsymbol{y_m}} \\ &= f_{M|Y}(\boldsymbol{m}|\boldsymbol{y_o}, \boldsymbol{\psi}) f(\boldsymbol{y_o}|\boldsymbol{\theta}) \end{aligned} \tag{2.16}$$

which is the factored likelihood from (2.15). Hence, this process is ignorable. When the process is MNAR we have

$$f(\boldsymbol{y_o}, \boldsymbol{m}|\boldsymbol{\theta}, \boldsymbol{\psi})d_{\boldsymbol{y_m}} = \int f_{M|Y}(\boldsymbol{m}|\boldsymbol{y_o}, \boldsymbol{y_m}, \boldsymbol{\theta}, \boldsymbol{\psi}) f(\boldsymbol{y_o}, \boldsymbol{y_m}|\boldsymbol{\theta}, \boldsymbol{\psi})d_{\boldsymbol{y_m}} \tag{2.17}$$

The final line of (2.17) can not be solved in the same way as (2.16) since the missing process is affected by the unobserved value y_m.

2.5.1 Pattern Mixture Models and Selection Models

Under the MNAR assumption, there are several proposed ways of modeling, including pattern mixture models and selection models [Little and Rubin, 2019], which are different factorizations of the full conditional (2.10). From now on, let $\boldsymbol{x_i}$ be the set of fully observed explanatory variables and ϵ_i be an unobserved within-subject random effect with hyperparameter γ. [Little and Rubin, 2019]. The pattern mixture model can be viewed as a mixture of different populations who are characterized by their missingness pattern [Gad and Darwish, 2013] and is defined as follows,

$$f(\boldsymbol{y_i}, \boldsymbol{m_i}, \epsilon_i|\boldsymbol{x_i}, \boldsymbol{\theta}, \boldsymbol{\psi}, \gamma) = f(\boldsymbol{y_i}|\boldsymbol{x_i}, \epsilon_i, \boldsymbol{m_i}, \boldsymbol{\theta}) f(\epsilon_i|\boldsymbol{x_i}, \boldsymbol{m_i}, \gamma) f(\boldsymbol{m_i}|\boldsymbol{x_i}, \boldsymbol{\psi}). \tag{2.18}$$

The selection model is based on the following factorization,

$$f(\boldsymbol{y_i}, \boldsymbol{m_i}, \epsilon_i|\boldsymbol{x_i}, \boldsymbol{\theta}, \boldsymbol{\psi}, \gamma) = f(\boldsymbol{m_i}|\epsilon_i, \boldsymbol{x_i}, \boldsymbol{y_i}, \boldsymbol{\psi}) f(\boldsymbol{y_i}|\epsilon_i, \boldsymbol{x_i}, \boldsymbol{\theta}) f(\epsilon_i|\boldsymbol{x_i}, \gamma). \tag{2.19}$$

The first factor models the missing mechanism, and the second the distribution of y_i in the population, and $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are distinct.

2.5.2 Shared Parameter Models

This work uses a class of selection models known as Shared Parameter Models (SPM) [Little and Rubin, 2019, Chap. 15.2] defined as follows:

$$g(\boldsymbol{y_i}, \boldsymbol{m_i}, \epsilon_i|\boldsymbol{x_i}, \boldsymbol{\theta}, \boldsymbol{\psi}, \gamma) = g(\boldsymbol{y_i}|\boldsymbol{x_i}, \epsilon_i, \boldsymbol{\theta}) g(\boldsymbol{m_i}|\boldsymbol{x_i}, \epsilon_i, \boldsymbol{\psi}) g(\epsilon_i|\boldsymbol{x_i}, \gamma). \tag{2.20}$$

This model assumes that the measurement and dropout processes depend on a shared latent variable ϵ_i. This allows dropouts to be classified as ignorable provided $g(m_i|\boldsymbol{x_i}, \epsilon_i, \boldsymbol{\psi}) = g(m_i|\boldsymbol{x_i}, \boldsymbol{\psi})$ [Vonesh et al., 2006].

2.6 Scoring Rules

Model comparison is important to obtain the best possible model fits. To compare models, the use of scoring rules is common. A scoring rule accesses the probability

forecast of a model fit by assigning a numerical score to the predictive distribution and the observed value [Gneiting and Raftery, 2007]. A scoring rule is strictly proper if there only exists one optimal model [Gneiting and Raftery, 2007]. I.e., it is impossible for two competing models to obtain the same score, thereby allowing for competing models to be ranked.

To evaluate the models produced in this work, we use the continuous ranked probability score (CRPS) [Brown, 1974, Hersbach, 2000, Krüger et al., 2020] for continuous responses and the Brier score [Brier et al., 1950] for binary responses. These are both strictly proper scoring rules, and both these scores penalize deviation from the true observed value. A lower score, therefore, indicates a better model.

Let $f(x)$ be the probability density function of some forecast and y is the actual observed value. The CRPS score is defined following the notation of Hersbach [2000];

$$CRPS(F, y) = \int_{-\infty}^{\infty} [F(x) - F_y(x)]^2 dx. \tag{2.21}$$

where F and F_y are cumulatively distributions given by, $F(x) = \int_{-\infty}^{x} f(y)dy$, and $F_y(x) = H(x - y)$ where

$$H(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geqslant 0, \end{cases}$$

is the Heaviside function. The empirical equivalent to Equation (2.21) reduces to,

$$CRPS(\hat{F}_k, y) = \frac{1}{k} \sum_{i=1}^{k} |X_i - y| - \frac{1}{2k^2} \sum_{i=1}^{k} \sum_{j=1}^{k} |X_i - X_j|, \tag{2.22}$$

where $\hat{F}_k$ is the empirical cumulative distribution of the k available simulated samples (Krüger et al. [2020], Jordan et al. [2017]) and $X_1, \ldots, X_k \sim F$ are the simulated samples from the distribution of interest, F. The function **crps** from the R package **scoringRules** (Jordan et al. [2017]) is used. In fact this package uses an equivalent representation of Equation 2.22 given by

$$CRPS(\hat{F}_k, y) = \frac{2}{k^2} \sum_{i=1}^{k} (X_{(i)} - y) \Big(kH(X_{(i)} - y) - i + 0.5 \Big),$$

which uses the ordered simulated samples. Further detail can be seen in Jordan [2016]. A lower CRPS value suggests a better model as this score penalizes deviation from the true observed value. To compare models we use the mean CRPS given by

$$\overline{CRPS}(\hat{\boldsymbol{F}}_{\boldsymbol{k}}, \boldsymbol{y}) = \frac{1}{n} \sum_{i=1}^{n} CRPS(\hat{F}_{ki}, y_i), \tag{2.23}$$

where i indicates subject i It is evidently only possible to compute the CRPS for non-missing data where we have observed observations.

The Brier score is suitable for binary responses and is defined as follows,

$$\text{Brier}(\boldsymbol{p}, \boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2.$$

Here p_i is the probability of success and m_i the actual observation. In this work we simulate k predictions of probability p_i and use the mean of these predictions as follows,

$$\text{Brier}(\boldsymbol{P}, \boldsymbol{m}) = \frac{1}{n}\sum_{i=1}^{n}(m_i - \frac{1}{k}\sum_{j=1}^{k}p_j)^2. \tag{2.24}$$

$\boldsymbol{P}$ is the matrix containing k predictions for every subject $i = 1, \ldots, n$ and $\boldsymbol{m}$ is a vector containing the observed values.

3 Case Study: A Blood Pressure Predictive Model based on the HUNT Study.

This work is inspired by constructing predictive models for future blood pressure (BP_F) based on age, sex, BMI, and initial blood pressure from the HUNT Study accounting for data MNAR. Here we introduce the HUNT study and the specific data used. We set up the models and introduce the validation schemes we use.

3.1 The HUNT Study and Data for Blood Pressure Models

The HUNT Study is a longitudinal population-based health survey in central Norway [Krokstad et al., 2013, Åsvold et al., 2021]. Every adult citizen in the now former county of Nord-Trøndelag was invited to participate in the first survey in 1984-86 (HUNT1). All adult residents in the screening area have since then been invited to clinical examinations and questionnaires in 1995-97 (HUNT2), 2006-08 (HUNT3), and 2017-19 (HUNT4) [Krokstad et al., 2013, Åsvold et al., 2021].

In this study, observations of systolic blood pressure (BP), age (age), body mass index (BMI) and sex (sex, 0 for females and 1 for males) are used based on the findings by Espeland [2020]. A description of the variables is given in Table 3.1. When needed, a subscript indicates the HUNT survey of the observations (BP_2 denotes BP observed at HUNT2).

Table 3.1.: Summary of variables.

	Variable	Description (Unit)
Explanatory variables	BP_I	Initial systolic blood pressure* (mmHg)
	age	Age* (years)
	BMI	Body Mass Index* (kg/m^2)
	sex	Female (0) or Male (1)
Response variables	BP_F	Future systolic blood pressure* (mmHg)
	m	BP_F is observed (0) or missing (1)

*In the model fitting all continuous variables are standardized.

Following Tobin et al. [2005] BP is adjusted by adding 15 mmHg for all participants who self-reported using BP medication. We define a modeling cohort (HUNT2 cohort) consisting of 60385 participants with observations of initial blood pressure ($BP_I = BP_2$), age, BMI and sex in HUNT2, together with future blood pressure $BP_F = BP_3$ in HUNT3 and a missing indicator m (1 if missing, 0 if present at HUNT3). 43.1% of the participants in HUNT2 are missing in HUNT3. Dropout in this work is equivalent to "lost to follow-up" in medical terms. The validation cohort (HUNT3 cohort) consists of 50807 participants with observations of $BP_I = BP_3$, age, BMI and sex from HUNT3, and $BP_F = BP_4$ from HUNT3 in addition to a missing indicator m (1 if missing, 0 if present at HUNT4). In the HUNT3 cohort, 33.3% drop out prior to HUNT4. In all following analyses, BP_F, BP_I, age, and BMI observations are standardized by the corresponding sample mean and standard deviation in the HUNT2 cohort.

3.1.1 The Trøndelag Health Study Protocol

The HUNT Study protocols are described in detail in Krokstad et al. [2013] and Åsvold et al. [2021]. Here, we briefly overview the performed data collection relevant to this work. Age and sex were extracted from the Norwegian Population Registry. Height and weight were measured after removing shoes and other heavy clothing. The BP was measured, by trained personnel, in a sitting position after two minutes of rest. Three measurements were taken, one minute apart, of which the mean of the second and third was used to report BP. To assess the current use of BP medication, self-reported questionnaires were used. In HUNT2, the current use of BP medication was captured in "Are you taking medication for high blood pressure?" [Never; Previously; Currently]. In HUNT3, the question was reformulated to "Do you take, or have you taken medication for high blood pressure?" [No; Yes]. Therefore we combine this question with the answer to "If you are currently taking medicine for high blood pressure, have you felt unwell/ had side effects from this medicine?". We assume only the participant who currently takes BP medicine answered this question. In HUNT4, the use of BP medicine was captured through the question "Do you currently use any prescription medication for high blood pressure?" [No; Yes].

3.2 Explanatory Analysis

Summaries of both the HUNT2 and HUNT3 cohort are given in Table 3.2. The variables are summarized for all participants (present and missing), the present participants, and the missing participants for the HUNT2 and HUNT3 cohorts respectively. In addition the smoothed empirical distribution of BP_I, (BP_2 and BP_3), age, and BMI are plotted in Figure 3.1 grouped on missing status. Both from Table 3.2 and Figure 3.1, we see that there is a difference, especially for age, between present and missing participants. We

Table 3.2.: The sample mean and standard deviation of BP_F, BP_I, age, and BMI and proportion of female/male participants in the HUNT2 and HUNT3 survey are displayed in the third column. The fourth and fifth columns display sample mean for the present and missing participants in the HUNT2 and HUNT3 cohort in addition to proportions of present/missing participants for the whole cohorts and per sex.

Summary of the HUNT2 cohort

Variable	Unit	HUNT2	Present in HUNT3	Missing in HUNT3
BP_3 (BP_F)	mmHg	-	136.1	-
BP_2 (BP_I)	mmHg	139.5 (23.6)	135.2	145.0
age_2	years	50.0 (17.1)	47.0	54.01
BMI_2	kg/m^2	26.4 (4.1)	26.2	26.6
sex			56.9 %	43.1 %
female	0	53.0 %	59.3 %	40.7 %
male	1	47.0 %	54.3 %	45.7 %

Summary of the HUNT3 cohort

Variable	Unit	HUNT3	Present in HUNT4	Missing in HUNT4,
BP$_4$ (BP_F)	mmHg	-	136.48	-
BP$_3$ (BP_I)	mmHg	133.21 (20.71)	131.48	136.67
age$_3$	years	53.08 (16.01)	51.68	55.90
BMI$_3$	kg/m^2	27.17 (4.41)	27.12	27.28
sex			66.7 %	33.3 %
female	0	54.6 %	68.7 %	31.3 %
male	1	45.4 %	64.3 %	35.7 %

also see that more males than females drop out of the study. These differences suggest that the data is at least MAR. Further examination is needed to explore the possibility of data MNAR. In addition, we see that middle-aged participants are less likely to drop out than young or elderly participants suggesting that age will affect the probability of dropping out in a non-linear way.

3.3 Shared Parameter Models for Blood Pressure Based on the HUNT Study

We set up shared parameter models for BP_F and the missing process using age, sex, BMI, and BP_I as explanatory variables. Let BP_{Fi} and m_i represent future blood pressure and missing status for individual i. We will set up SPMs in the framework of a latent Gaussian model as presented in Section 2.2. The likelihoods are chosen to be Gaussian for BP_F and Bernoulli with logit link for m_i.

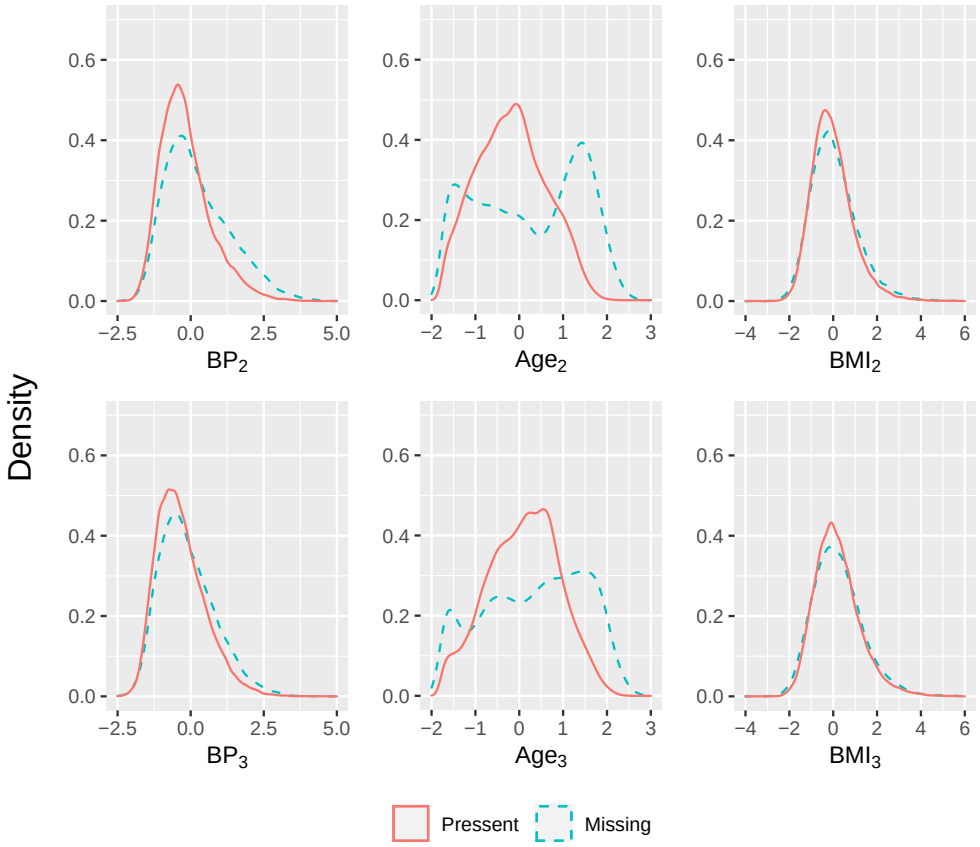


Figure 3.1.: Smoothed empirical density of BP_I (in the HUNT2 cohort $BP_I = BP_2$ and in the HUNT3 cohort $BP_I = BP_3$), age, and BMI for all participants from the HUNT2 (upper panel) and HUNT3 (lower panel) cohort grouped by missing status. All the variables are displayed in standardized values.

3.3.1 Shared Parameter Model, a General Formulation

This work uses a variation of the SPM defined in (2.20). Let $Y = (\boldsymbol{y}, \boldsymbol{m})^T$ denote a two dimensional response vector consisting of measurement $\boldsymbol{y}$ and missing status $\boldsymbol{m}$. Let X the $n * m$ matrix of explanatory variables, β a vector of linear effects, $\boldsymbol{f}(X)$ a vector of non-linear effects, and $\boldsymbol{\epsilon}$ a vector of individual random effects, $\epsilon_i \sim N(0, \sigma)$. The general form of the SPM used in this work is then given by,

$$\boldsymbol{\eta_Y} = X\boldsymbol{\beta} + \boldsymbol{f}(X^T) + (\boldsymbol{\epsilon}, c\boldsymbol{\epsilon})^T I$$

where I is the identity matrix with shape $n * n$. $\boldsymbol{\eta_Y}$ is connected to the expectation $E(Y) = \boldsymbol{\mu}$ through link functions.

This can also be written as two submodels, one for the measurement process and one for the missing process, respectively,

$$\begin{aligned}\boldsymbol{\eta_y} &= \boldsymbol{X}\boldsymbol{\beta_y} + \boldsymbol{f_y}(\boldsymbol{X}) + \boldsymbol{\epsilon},\\ \boldsymbol{\eta_m} &= \boldsymbol{X}\boldsymbol{\beta_m} + \boldsymbol{f_m}(\boldsymbol{X}) + c\boldsymbol{\epsilon},\end{aligned}$$

linked together by the shared random effect $\boldsymbol{\epsilon}$.

The measurement process has a binary outcome. Therefor, $\boldsymbol{\eta_m}$ is linked to the expectation of $\boldsymbol{m}$ through a logit link function. The measurement process can be any distribution belonging to the exponential family. Thus, $\boldsymbol{\eta_y}$ can be linked to the expectation of $\boldsymbol{y}$ in a number of ways. In this work however, the measurement process follows a latent Gaussian distribution, and we use the identity link function. This means,

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta_y} + \boldsymbol{f_y}(\boldsymbol{X}) + \boldsymbol{\epsilon},\\ \boldsymbol{\epsilon} &\sim N(0, \sigma),\\ logit(\boldsymbol{p}) &= \boldsymbol{X}\boldsymbol{\beta_m} + \boldsymbol{f_m}(\boldsymbol{X}) + c\boldsymbol{\epsilon},\\ \boldsymbol{m} &\sim Bernoulli(\boldsymbol{p}).\end{aligned} \tag{3.1}$$

3.3.2 Shared Parameter Model With Additive Effects

To explore the need for non-linear effects, we first model all continuous variables as additive effects ($f(z)$) through a random walk of order 2 with a sum to zero constraints [Gómez-Rubio, 2020]. Using a second-order random walk as an additive effect ensures a Gaussian latent field, necessary to use INLA.

The random walk model of order 2 assumes independent second order increments. Let $m = 1 : n$ be the index for the increments and define

$$\Delta^2 z_m = z_m - 2z_{m+1} + z_{m+2} \sim N(0, \sigma_z^2).$$

The density for $f(\boldsymbol{z})$ is,

$$f(\boldsymbol{z}|\sigma_z) \propto \sigma_z^{-\frac{n-2}{2}} exp\left\{ -\frac{1}{2\sigma_z} \sum_{m=1}^{n-2} (\Delta^2 z_m) \right\}$$

The sum to zero constraint means the sum of all random effect components must be zero. For more information about random walks, see Rue and Held [2005].

The specification of the general SPM given in Equation (3.1) becomes,

$$\begin{aligned} BP_{Fi} = \alpha_0 + f_{BP_F}(BP_{Ii}) + f_{BP_F}(age_i) + f_{BP_F}(BMI_i) + \alpha_{sex} sex_i + \epsilon_i \\ \epsilon_i \sim N(0, \sigma_\epsilon^2) \\ \text{logit}(p_i) = \beta_0 + f_m(BP_{Ii}) + f_m(age_i) + f_m(BMI_i) + \beta_{sex} sex_i + c\epsilon_i \\ m_i \sim \text{Bernoulli}(p_i). \end{aligned} \tag{3.2}$$

Here, α_0, α_{sex}, β_0, and β_{sex} are the model parameters. For which we chose the following independent weak prior,

$$\alpha_0, \alpha_{sex}, \beta_0, \beta_{sex} \sim N(0, 10^{3^2}).$$

The individual random effect ϵ_i is normally distributed with mean zero and standard deviation σ_ϵ. Both the standard deviation of the additive effects $f(.)$ and ϵ_i have hyperparameters with the following independent prior distributions,

$$\begin{aligned} \sigma_{BP_F BP_I}, \sigma_{BP_F age}, \sigma_{BP_F BMI}, \sigma_{mBP_I}, \sigma_{mage}, \\ \sigma_{mBMI}, \sigma_\epsilon \sim \text{Gamma}(1, 5 \cdot 10^5). \end{aligned}$$

The subscript to σ indicates which additive effect the variance belongs to.

The association parameter is given the following prior.

$$c \sim N(0, 1^2),$$

The association parameter c is expected to be no more than order one. This expectation becomes clear when we observe the results of the naive model introduced in Section 3.4 given in Table 4.1. For the naive model $\sigma_\epsilon = 0.77$ (Table 4.1). Hence, ϵ_i takes values of the same order as the standardized explanatory variables. We do not expect the SPM to deviate drastically from the naive model. Therefore we expect ϵ_i to be of a similar

order for the SPM as for the naive model and affect the dropout process similarly as the other explanatory variables. For ϵ_i to affect the dropout process similarly as the other variables, the association parameter must be of the same order as the other parameters in the latent field. These are all less than one.

3.3.3 Shared Parameter Model for Blood Pressure and Missing Status

Based on the results presented in Section 4.1 we decide to only model the age effect in the missing process additively.

This leads to the following submodels for predicting BP_{Fi} and m_i for participant i:

$$\begin{aligned} BP_{Fi} &= \alpha_0 + \alpha_{BP} BP_{Ii} + \alpha_{age} age_i + \alpha_{BMI} BMI_i + \alpha_{sex} sex_i + \epsilon_i \\ \epsilon_i &\sim N(0, \sigma_\epsilon^2) \\ \text{logit}(p_i) &= \beta_0 + \beta_{BP} BP_{Ii} + f(age_i) + \beta_{BMI} BMI_i + \beta_{sex} sex_i + c\epsilon_i \\ m_i &\sim \text{Bernoulli}(p_i). \end{aligned} \tag{3.3}$$

This model is from now on referred to as the SPM and is our main model.

The individual random effect is Gaussian distributed with zero mean and variance one. Here, α_0, α_{BP}, α_{age}, α_{BMI}, α_{sex}, β_0, β_{BP}, β_{BMI}, and β_{sex} are the model parameters with the following independent priors,

$$\alpha_0, \alpha_{BP}, \alpha_{age}, \alpha_{BMI}, \alpha_{sex}, \beta_0, \beta_{BP}, \beta_{BMI}, \beta_{sex} \sim N(0, 10^{3^2}).$$

The prior distributions for σ_{age} and σ_ϵ is given by,

$$\sigma_{age}, \sigma_\epsilon \sim \text{Gamma}(1, 5 \cdot 10^5).$$

The association parameter c is given the following normal prior,

$$c \sim N(0, 1^2),$$

for the same reason as given in Section 3.3.2.

3.4 Naive Model for Blood Pressure and Missing Status

We compare the SPM with a naive model assuming data to be MAR. The naive model is identical to the model specified in (3.3) except that the individual random effect is no longer shared, meaning $c = 0$. The formulation becomes,

$$\begin{aligned} BP_{Fi} &= \alpha_0 + \alpha_{BP} BP_{Ii} + \alpha_{age} age_i + \alpha_{BMI} BMI_i + \alpha_{sex} sex_i + \epsilon_i \\ \epsilon_i &\sim N(0, \sigma_\epsilon) \\ \text{logit}(p_i) &= \beta_0 + \beta_{BP} BP_i + f(age_i) + \beta_{BMI} BMI_i + \beta_{sex} sex_i \\ m_i &\sim \text{Bernoulli}(p_i). \end{aligned} \tag{3.4}$$

All model parameters in the naive model have identical priors as stated for the SPM (3.3). Even though the naive model consists of two independent models, we refer to both submodels together as the naive model and use it as a benchmark model.

3.5 Inference for Predictive Distributions

For simplicity later on we introduce some notation for the variables and parameters introduced in Section 3.3.2, Section 3.3.3, and Section 3.4. Let $\boldsymbol{x_i} = (BP_{Ii}, age_i, BMI_i, sex_i)$ be the set of explanatory variables and let $\boldsymbol{y_i} = (BP_{Fi}, m_i)$ be the response variables. Further let $X = (x_i, .., x_n)^T$ be the fully observed dataset of all explanatory variables for n participants and $Y = (y_1, ..., y_n)^T$ be the corresponding response values. When needed we use a subscript to indicate respective HUNT surveys. Denote the latent field by $\boldsymbol{\theta} = (\alpha_0, \alpha_{BP}, \alpha_{age}, \alpha_{BMI}, \alpha_{sex}, \beta_0, \beta_{BP}, \beta_{BMI}, \beta_{sex}, \epsilon, f, c)$. f denotes the additive effect for age. Let the parameters of the latent field associated to the blood pressure model be denoted by $\boldsymbol{\theta_{BP}} = (\alpha_0, \alpha_{BP}, \alpha_{age}, \alpha_{BMI}, \alpha_{sex}, \epsilon)$ and the latent field associated to the missing process be denoted by $\boldsymbol{\theta_m} = (\beta_0, \beta_{BP}, \beta_{BMI}, \beta_{sex}, c, \epsilon, f)$. The set of non-Gaussian hyper parameters are denoted $\boldsymbol{\gamma} = (\sigma_\epsilon, \sigma_{age})$.

When fitting the SPM and the naive model we obtain the posterior distribution of the latent field $\pi(\boldsymbol{\theta}|X, Y)$ and the hyperparameters $\pi(\boldsymbol{\gamma}|X, Y)$. The posterior distribution of θ_k is given by $\pi(\theta_k|X, Y) = \int \pi(\boldsymbol{\theta}|X, Y) d\theta_{-k}$ for all k model parameters. $d\theta_{-k}$ denotes the integration of all parameters but θ_k.

We can use the posterior distributions of the parameters to obtain the predictive distribution of $\boldsymbol{y_{new}} = (BP_{Fnew}, m_{mew})$ for a new observation $\boldsymbol{x_{new}}$. In mathematical terms,

$$\begin{aligned}\pi(\boldsymbol{y_{new}}|\boldsymbol{x_{new}}, X, Y) &= \int \pi(\boldsymbol{y_{new}}, \boldsymbol{\theta}|\boldsymbol{x_{new}}, X, Y)d\boldsymbol{\theta} \\ &= \int \pi(\boldsymbol{y_{new}}|\boldsymbol{x_{new}}, X, Y, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{x_{new}}, X, Y)d\boldsymbol{\theta}\end{aligned}$$

which can be simplified to

$$\pi(\boldsymbol{y_{new}}|\boldsymbol{x_{new}}, X, Y) = \int \pi(\boldsymbol{y_{new}}|\boldsymbol{x_{new}}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|X, Y)d\boldsymbol{\theta}$$

Similarly this gives

$$\pi(BP_{Fnew}|\boldsymbol{x_{new}}, X, Y) = \int \pi(BP_{Fnew}|\boldsymbol{x_{new}}, \boldsymbol{\theta_{BP}})\pi(\boldsymbol{\theta_{BP}}|X, Y)d\boldsymbol{\theta_{BP}}$$

and

$$\pi(m_{new}|\boldsymbol{x_{new}}, X, Y) = \int \pi(m_{new}|\boldsymbol{x_{new}}, \boldsymbol{\theta_m})\pi(\boldsymbol{\theta_m}|X, Y)d\boldsymbol{\theta_m}$$

for the posterior predictive distribution for a new value of BP_F and m. For simplicity we compute and evaluate p_{new} instead of m_{new} where $m_{new} \sim Bernoulli(p_{new})$.

3.6 Parameter Estimation

We fit both the SPMs (3.2), (3.3) and the naive model (3.4) using the HUNT2 cohort, obtaining posterior distributions $\pi(\boldsymbol{\theta}|X_2, Y_2)$ and $\pi(\boldsymbol{\gamma}|X_2, Y_2)$ where the subscript 2 indicates the HUNT2 cohort. These posteriors are later used to obtain predictive distributions used for validation of model predictions and to check if the data is MNAR.

For completeness we also fit the SPM (3.3) and naive model (3.4) to the HUNT3 cohort obtaining $\pi(\boldsymbol{\theta}|X_3, Y_3)$ and $\pi(\boldsymbol{\gamma}|X_3, Y_3)$ where the subscript 3 indicates the HUNT3 cohort.

3.7 Validation of the Model Predictions of Blood Pressure

To validate the model predictions we obtain the posterior predictive distribution for future blood pressure (BP_F) of all participants in the HUNT3 cohort. This means we obtain,

$$\pi(Y_3|X_3, Y_2, X_2) = \int \pi(Y_3|X_3, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|X_2, Y_2)d\boldsymbol{\theta} \tag{3.5}$$

We evaluate the performance of the SPM and naive model using the mean CRPS score ($\overline{CRPS}$) for the posterior predictive distributions of $\boldsymbol{BP_F}$ and the Brier score for the posterior mean prediction of $\boldsymbol{p}$. The blood pressure model can only be validated on the participants observed in both the HUNT3 and the HUNT4 surveys. In contrast, the probability of dropping $\boldsymbol{p}$ out can be evaluated for all participants. We compare the posterior mean predictions of $\boldsymbol{p}$ for the SPM and naive model grouped on missing status in the HUNT3 cohort.

3.8 Evaluation of the Missing Not at Random Assumption Based on Conditioning on the Missing Process

We suggest a novel method for validating if data is MNAR if the data follows the SPM, by obtaining the posterior predictive distributions $\pi(\boldsymbol{BP_F}|\boldsymbol{m})$ and $\pi(\boldsymbol{BP_F})$ for the HUNT3 cohort and comparing the posterior predictive means to the observed $\boldsymbol{BP_F}$ values. We denote the posterior predictive distributions $\pi(\boldsymbol{BP_F}|\boldsymbol{m})$ and $\pi(\boldsymbol{BP_F})$, $\boldsymbol{\hat{BP}_F}|\boldsymbol{m}$ and $\boldsymbol{\hat{BP}_F}$ respectively. If the data is MNAR, there is a connection between the dropout process and the blood pressure process. Knowledge of the missing status of a participant i should then contain information about the shared individual random effect ϵ_i. Since ϵ_i is shared between the two sub-models having this information about ϵ_i should improve $\hat{BP}_{Fi}$. Therefore, we expect $\hat{BP}_{Fi}|m$ to be better than $\hat{BP}_{Fi}$ if the data are MNAR and follows the SPM. If the data is MAR, we expect the predictions $\hat{BP}_{Fi}|m$ and $\hat{BP}_{Fi}$ to be equally good. Formally this means we compare,

$$\begin{aligned} \hat{BP}_{Fi}|m_i = \pi(BP_{Fi}|m_i) &= \int \pi(BP_{Fi}, \epsilon_i|m_i)d\epsilon_i \\ &= \int \pi(BP_{Fi}|m_i, \epsilon_i)\pi(\epsilon_i|m_i)d\epsilon_i \end{aligned}$$

with

$$\begin{aligned} \hat{BP}_{Fi} = \pi(BP_{Fi}) &= \int\int \pi(BP_{Fi}, \epsilon_i, m_i)d\epsilon_i dm_i \\ &= \int\int \pi(BP_{Fi}, m_i|\epsilon_i)\pi(\epsilon_i)d\epsilon_i dm_i \\ &= \int\int \pi(BP_{Fi}|\epsilon_i, m_i)\pi(m_i|\epsilon_i)\pi(\epsilon_i)d\epsilon_i dm_i \end{aligned}$$

For simplicity, we use the posterior mean of $\hat{BP}_{Fi}|m$ and $\hat{BP}_{Fi}$, which we refer to with the same notation, and compute the mean absolute error (MAE) of $\boldsymbol{\hat{BP}_F|m}$ and $\boldsymbol{\hat{BP}_F}$ for the participants present in the follow-up survey.

3.9 Software Implementation

In this work, we use the R-INLA software [R-INLA, 2021]. This software provides a relatively easy way to use INLA and is used by many authors for spatial statistics [Blangiardo et al., 2013, Lindgren and Rue, 2015, Blangiardo and Cameletti, 2015, Bakka et al., 2018]. However, this software has many more applications [Martins et al., 2013, Gómez-Rubio, 2020, Niekerk et al., 2021] and is also previously used to fit SPMs [Steinsland et al., 2014, Espeland, 2020].

Documentation regarding the R-INLA software can be found at the R-INLA homepage [R-INLA, 2021]. The R-INLA software supports fitting models with multiple likelihoods [Steinsland et al., 2014, Espeland, 2020, Gómez-Rubio, 2020, Chap. 6.4] which is the case for the SPM (3.3). An easy-to-follow example of implementing multiple likelihoods in R-INLA can be found in [Gómez-Rubio, 2020, Chap. 6.4].

The two main functions from the R-INLA library used in this work are inla(), providing marginal posterior distributions for the model parameters, and inla.posterior.sample() to sample from the posterior distributions of the model parameters.

All the code used to fit the models is available in the GitHub repository by Hofman [2021a].

3.9.1 Shared Parameter Model Fit

When implementing the SPMs (3.2) and (3.3), introduced in Section 3.3.2 and Section 3.3.3, in INLA, for technical reasons we have to rewrite the models slightly by adding $\sigma_{2_{\epsilon_2}} = 0.001$. The two models are then given by:

$$\begin{aligned} \mu_i &= \alpha_0 + f_{BP_F}(BP_{Ii}) + f_{BP_F}(age_i) + f_{BP_F}(BMI_i) + \alpha_{sex} sex_i + \epsilon_i \\ BP_{Fi} &\sim N(\mu_i, \sigma_{2_{\epsilon_2}}) \\ \text{logit}(p_i) &= \beta_0 + f_m(BP_{Ii}) + f_m(age_i) + f_m(BMI_i) + \beta_{sex} sex_i + c\epsilon_i \\ m_i &\sim \text{Bernoulli}(p_i), \end{aligned} \tag{3.6}$$

and

$$\begin{aligned}
\mu_i &= \alpha_0 + \alpha_{BP} BP_{Ii} + \alpha_{age} age_i + \alpha_{BMI} BMI_i + \alpha_{sex} sex_i + \epsilon_i \\
BP_{Fi} &\sim N(\mu_i, \sigma_{2_{\epsilon_2}}) \\
\text{logit}(p_i) &= \beta_0 + \beta_{BP} BP_{Ii} + f(age_i) + \beta_{BMI} BMI_i + \beta_{sex} sex_i + c\epsilon_i \\
m_i &\sim \text{Bernoulli}(p_i).
\end{aligned} \tag{3.7}$$

The priors for all model parameters of (3.6) and (3.7) are the same as given in Section 3.3.2 and Section 3.3.3 respectively. Since $\sigma_{2_{\epsilon_2}}$ is very small we ensures all the variance is still captured by ϵ_i and the random effect can be shared between the two models. The Gaussian latent field in (3.6) consists of $\boldsymbol{\theta} = (\alpha_0, f_{BP_F}(.), \alpha_{sex}, \epsilon_i, \beta_0, f_m(.), \beta_{sex}, c)$ where $f_{BP_F}(.)$ and $f_m(.)$ are all the non-linear effects. The non-Gaussian hyperparameters in (3.6) are $\boldsymbol{\gamma} = (\sigma_{BP_F BP_I}, \sigma_{BP_F age}, \sigma_{BP_F BMI}, \sigma_{mBP_I}, \sigma_{mage}, \sigma_{mBMI}, \sigma_\epsilon)$.

In our main SPM (3.7) the Gaussian latent field is given by $\boldsymbol{\theta} = (\alpha_0, \alpha_{BP}, \alpha_{age}, \alpha_{BMI}, \alpha_{sex}, \epsilon_i, \beta_0, \beta_{BP}, f(age), \beta_{BMI}, \beta_{sex}, c)$ and the non-Gaussian hyperparameters are $(\boldsymbol{\gamma} = \sigma_{age}, \sigma_\epsilon)$.

3.9.2 Validation of the Model Predictions of Blood Pressure

The validation scheme introduced in Section 3.7 is implemented as follows: To obtain the posterior predictive distribution for individual i, $\pi(BP_{Fi})$ and $\pi(p_i)$, we use $j = 1:300$ samples from the marginal posterior distributions of the latent field

$$\pi(\boldsymbol{\theta}_{true}|X_2, Y_2)) = \pi(\hat{\alpha}_0, \hat{\alpha}_{BP}, \hat{\alpha}_{age}, \hat{\alpha}_{BMI}, \hat{\alpha}_{sex}, \hat{\beta}_0, \hat{\beta}_{BP}, \hat{\beta}_{BMI}, \hat{\beta}_{sex}, \hat{c}, \hat{\sigma}_\epsilon),$$

from both the SPM and naive model presented in Section 3.3 and 3.4. This is done with the help of the function inla.posterior.sample(). The age effect is taken from the SPM (3.3) and naive model (3.4), respectively. For each sample j we compute the corresponding values of BP_{Fij} and p_{ij} as follows:

$$\begin{aligned}
\epsilon_{ij} &\sim N(0, \hat{\sigma}_{\epsilon_j}) \\
BP_{Fij} &= \hat{\alpha}_{0_j} + \hat{\alpha}_{BP_j} BP_{3i} + \hat{\alpha}_{age_j} age_{3i} + \hat{\alpha}_{BMI_j} BMI_{3i} + \hat{\alpha}_{sex_j} sex_{3i} + \epsilon_{ij} \\
\text{logit}(p_{ij}) &= \hat{\beta}_{0_j} + \hat{\beta}_{BP_j} BP_{3i} + \hat{f}(age_{3i}) + \hat{\beta}_{BMI_j} BMI_{3i} + \hat{\beta}_{sex_j} sex_{3i} + \hat{c}\epsilon_{ij}
\end{aligned} \tag{3.8}$$

3.9.3 Evaluation of the Missing Not at Random Assumption Based on Conditioning on The Missing Process

We implement the validation scheme presented in Section 3.8 to evaluate the MNAR assumption in the following way.

When predicting $\boldsymbol{B\hat{P}_F}|\boldsymbol{m}$, the observed $\boldsymbol{BP_F}$ values are removed for all participants. Then we refit the SPM (3.3) with normal priors using the posterior mean estimates $\hat{\boldsymbol{\theta}}$ of the model parameters, obtained from fitting the SPM (3.3) presented in Section 3.3, as mean and a very small variance $\sigma = 0.001$ meaning $\theta_k \sim N(\hat{\theta}_k, 0.001^2)$. INLA fits missing values from its marginal posterior distribution using the posterior mean. When predicting $\boldsymbol{B\hat{P}_F}$, both $\boldsymbol{BP_F}$ and $\boldsymbol{m}$ are removed from the data before we refit the SPM as for $\boldsymbol{B\hat{P}_F}|\boldsymbol{m}$. The missing values are fitted with the posterior mean of the posterior marginal predictive distribution of BP_{Fi} and m_i. Hence, we obtain the posterior mean estimate of $BP_{Fi}|m_i$ and BP_{Fi}, $B\hat{P}_{Fi}|m$ and $B\hat{P}_{Fi}$, for all participants. We compute the MAE for the participants present at HUNT4.

4 Results from the Blood Pressure Case Study

The two shared parameter models (SPM), (3.2) and (3.3), and the naive model (3.4) introduced in Chapter 3 are fitted to the HUNT2 cohort as described in Section 3.3 and Section 3.4. This chapter presents the resulting parameter estimates for the SPM with additive effects for all continuous variables (3.2) and arguments for the removal of some of these additive effects leading to our main SPM. Further, we present the parameter estimates of the model fits of the main SPM (3.3) and the naive model (3.4) on the HUNT2 cohort. We briefly explore the same model fits on the HUNT3 cohort and compare the posterior mean estimates obtained on the HUNT2 and HUNT3 cohort, respectively. Then we validate the predictive models obtained based on the HUNT2 cohort by predicting BP_F and missing status in the HUNT3 cohort as introduced in Section 3.7. Further, we test the novel validation scheme for evaluating if the data is MNAR described in Section 3.8 on the HUNT3 cohort.

4.1 Shared Parameter Model With Additive Effects

To explore the need for non-linear effects, we modeled the SPM with all continuous variables as a random walk of order two as presented in (3.2). Modeling all parameters in an additive way is extremely computationally demanding. With 32 CPU cores and 32 G memory, we were still only able to fit the SPM (3.2) with 15000 ($\approx 25\%$) participants. These participants were drawn randomly. Even though we were not able to fit the model on all participants, Figure 4.1 shows that the age effect in the missing process is non-linear. The other continuous variables, although not perfectly linear, are much closer to being linear even when we allow them not to be. Therefore we chose to model all variables linearly except for age in the dropout process, which we model as an additive effect. The final model can be seen in (3.3) in Section 3.3.3 and is referred to by either the main SPM or the SPM. This choice also means the computational time and resources needed to fit the model gets reduced significantly. If this model is to be used in future research, the model must be computationally feasible.

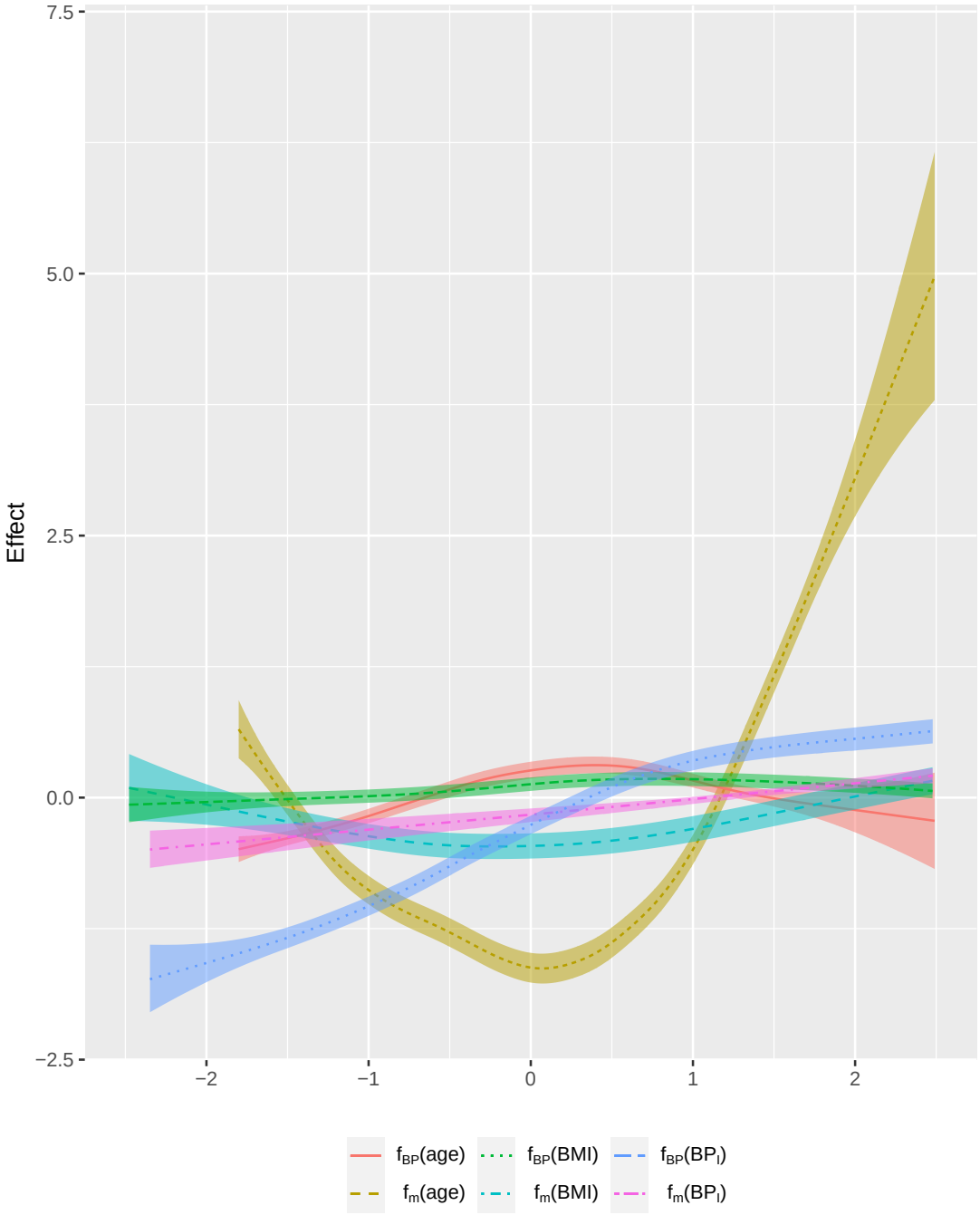


Figure 4.1.: Additive effects of age, BMI, and BP_I in the SPM for both the BP process and the missing process.

4.2 Shared Parameter and Naive Model Results for Blood Pressure and Missing Status on the HUNT2 Cohort

The posterior distributions of the estimates obtained by the SPM (3.3) and the naive model (3.4), introduced in Section 3.3.3 and 3.4 fitted to the HUNT2 cohort can be seen in Figure 4.2. The posterior mean and 95% credible intervals are presented in Table 4.1.

For the blood pressure submodel in the SPM, we see that the effect of BP_I is the largest, followed by age, BMI, and sex. The effect of BMI and sex are close to zero. The parameter estimates of the naive blood pressure model are of the same order as the SPM. However, in the blood pressure model, all variables but age have weaker effects in the naive model than for the SPM. The difference is especially pronounced for α_0 and α_{BP}, suggesting the two models could result in different model predictions.

For the missing process in the SPM, we see that sex has the largest effect on the probability of dropping out, followed by BP_I and BMI. The age effect is the largest for the elderly and the smallest for middle-aged participants Figure 4.3. The naive missing process again has the same order of variable effects. The SPM and the naive model are more similar in parameter estimates for the missing process than the blood pressure process. However, the naive parameter estimates are still shifted towards lower values for the naive model than the SPM.

The association parameter c, connecting the two submodels (3.3) is clearly positive, which implies an increase in the probability of dropping out for a larger random effect and a clear connection between the two sub-models.

Since the data used in this work contains personal information, we consider three female simulated participants with BP_I, age, and BMI, given in Table 4.2, to explore results on an individual level. We used a young and underweight female with low BP_I, a middle-aged and overweight female with normal BP_I, and an old and obese female with severely high BP_I. For these participants, the effect of the association parameter on the probability of dropping out is plotted in Figure 4.4 with individual random effects between -1.5 and 1.5, which corresponds to approximately two standard deviations of the random effect. We see an increase in the probability of dropping out for increased residuals. According to the SPM, this increase means that participants with higher BP_F than can be explained by the explanatory variables are more likely to drop out. When the random effect impacts the probability of dropping out, there is a connection between the submodels of the SPM, which indicates data MNAR.

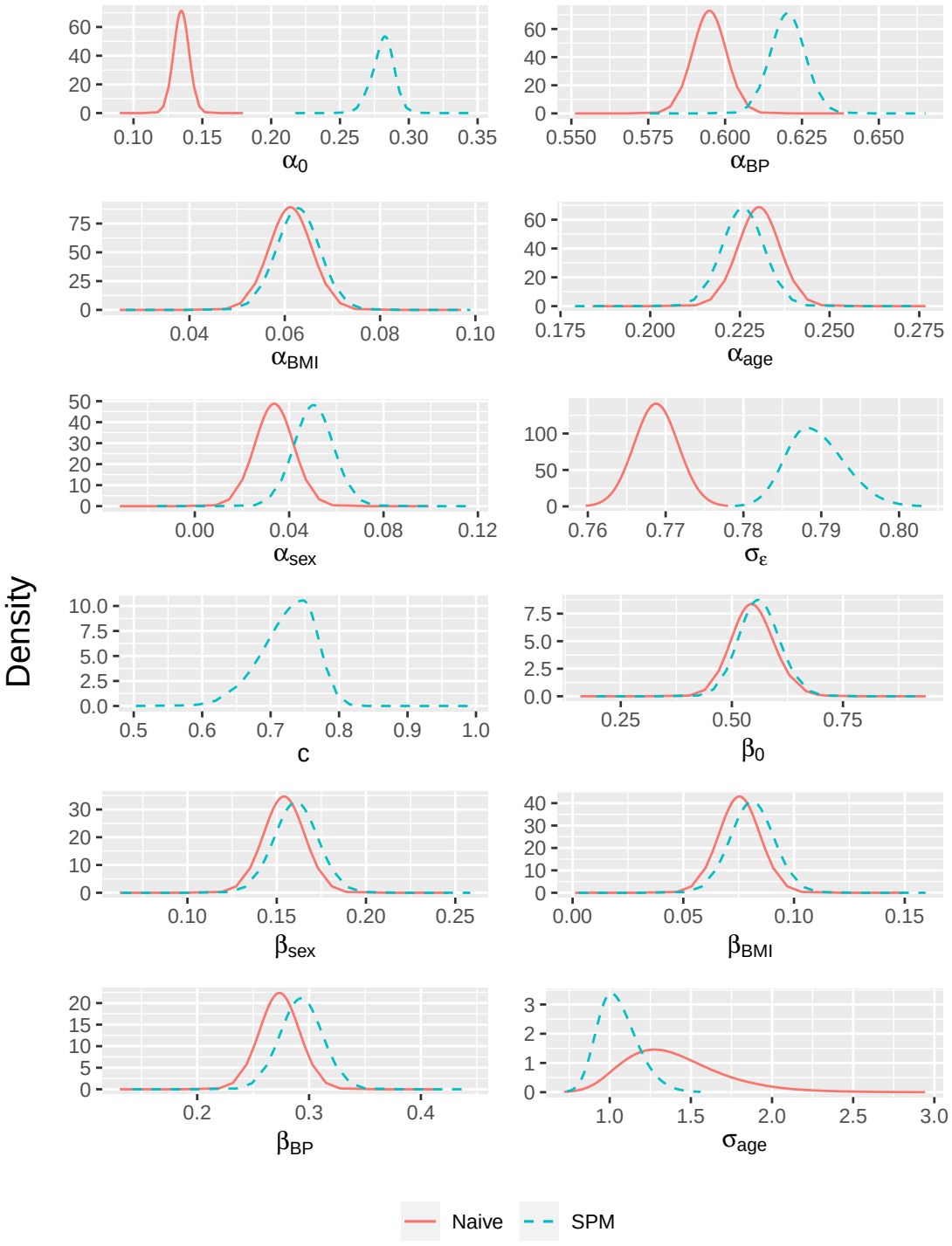


Figure 4.2.: Posterior distribution of the latent field and hyperparameters for the SPM (3.3) and naive model (3.4) fitted to the HUNT2 cohort.

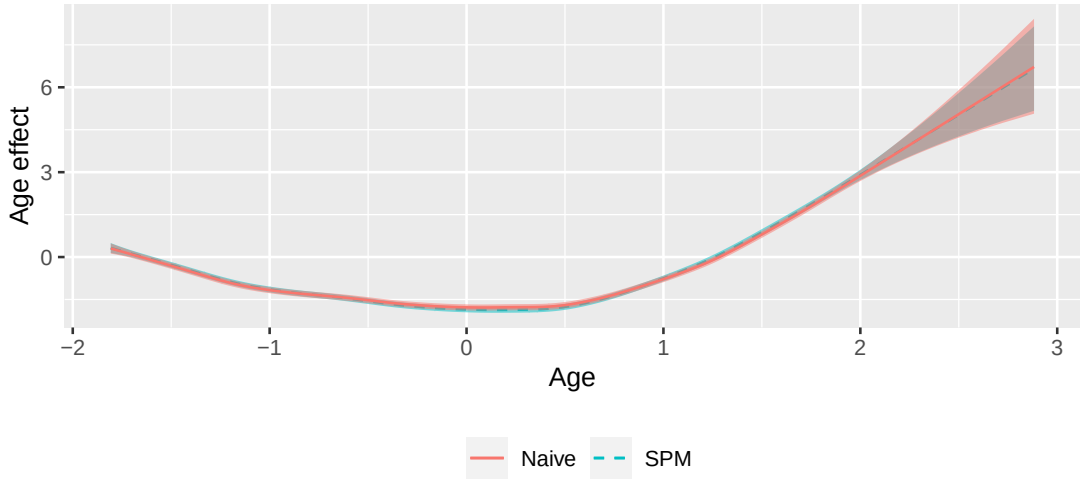


Figure 4.3.: Age effect for the SPM and naive model fitted to the HUNT2 cohort with 95% credible bands fitted to the HUNT2 cohort.

Table 4.1.: Summary of all parameters estimated for both the SPM and the naive model described in Section 3.3.3 an Section 3.4 fitted on the HUNT2 cohort. The posterior means and 95% credible intervals are displayed.

	SPM		**Naive**	
	Posterior mean	CI	Posterior mean	CI
α_0	0.275	(0.259, 0.291)	0.134	(0.123, 0.145)
α_{age}	0.244	(0.232, 0.255)	0.246	(0.235, 0.258)
α_{BMI}	0.073	(0.064, 0.082)	0.071	(0.062, 0.080)
α_{BP}	0.599	(0.589, 0.610)	0.578	(0.567, 0.588)
α_{sex}	0.041	(0.025, 0.057)	-0.022	(-0.064, 0.019)
β_0	0.567	(0.475, 0.666)	0.55	(0.460, 0.651)
β_{BMI}	0.087	(0.087, 0.106)	0.081	(0.063, 0.099)
β_{BP}	0.139	(0.139, 0.162)	0.133	(0.111, 0.155)
β_{sex}	0.292	(0.292, 0.329)	0.275	(0.240, 0.310)
σ_{age}	1.412	(0.946, 2.130)	1.400	(0.927, 2.123)
σ_ϵ	0.790	(0.783, 0.797)	0.77	(0.765, 0.776)
c	0.705	(0.645, 0.765)	-	-

Table 4.2.: Values of BP_I, age and BMI for three simulated female individuals.

id	BP_I^*	age^*	BMI^*	sex	BP_{Itrue}	age_{true}	BMI_{true}
1	-2	-1.5	-2	female	92.2	24.4	18.2
2	0	0	0	female	139.5	50.0	26.4
3	2	1.5	2	female	186.7	75.7	34.6

* Standardized values

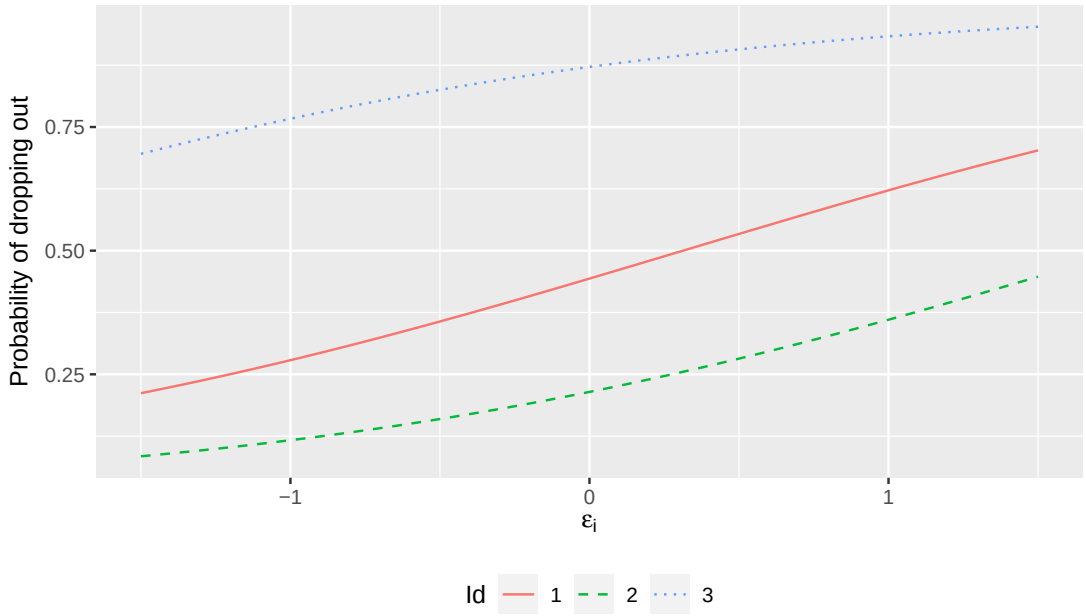


Figure 4.4.: The probability of dropping out as a function of the individual random effect of the blood pressure model (3.3) for three simulated female participants with BP_I, age, and BMI as given in Table 4.2.

4.3 Shared Parameter and Naive Model for Blood Pressure and Missing Status on the HUNT3 Cohort

For completeness we also fitted the SPM (3.3) and the naive model (3.4), introduced in Section 3.3.3 and Section 3.4 on the HUNT3 cohort.

We see from Figure 4.2 and Figure 4.5 that the relation between the SPM and naive model is similar for the HUNT3 cohort as for the HUNT2 cohort (Figure 4.2). We also see from Table 4.3 that the difference in parameter estimates are quite similar < 0.07 in absolute value, for all model parameters but β_0 and σ_{age}. Some difference is expected when working on real data. The HUNT surveys take place approximately 11 years apart. Hence, some societal changes affecting the general blood pressure and missing processes are inevitable.

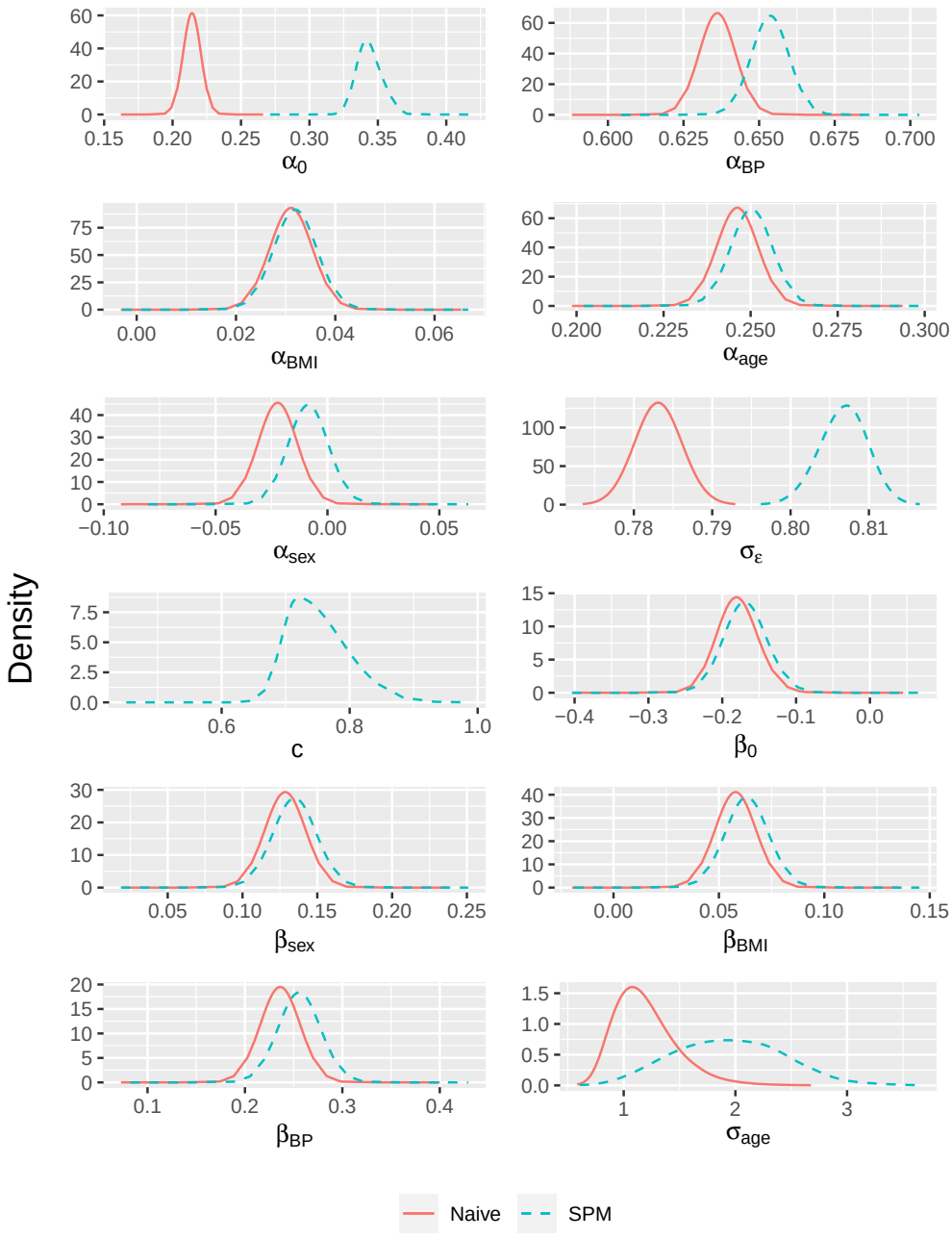


Figure 4.5.: Posterior distribution of the latent field and hyperparameters for the SPM (3.3) and naive model (3.4) fitted to the HUNT3 cohort.

Table 4.3.: The posterior mean of all model parameters, estimated for the SPM and the naive model on the HUNT2 and HUNT3 cohort.

	SPM Posterior Mean			**Naive** Posterior Mean		
	HUNT2 cohort	HUNT3 cohort	$Diff*$	HUNT2 cohort	HUNT3 cohort	$Diff*$
α_0	0.275	0.344	-0.063	0.134	0.214	-0.070
α_{age}	0.244	0.250	-0.024	0.246	0.246	-0.016
α_{BMI}	0.073	0.032	- 0.031	0.071	0.031	0.030
α_{BP}	0.599	0.654	- 0.037	0.578	0.636	-0.041
α_{sex}	0.041	-0.009	0.059	-0.022	-0.022	0.056
β_0	0.567	-0.168	0.73	0.55	-0.179	0.727
β_{BMI}	0.087	0.063	0.018	0.081	0.058	0.017
β_{BP}	0.139	0.135	0.026	0.133	0.129	0.026
β_{sex}	0.292	0.256	0.037	0.275	0.236	0.037
σ_{age}	1.412	1.927	-0.892	1.400	1.196	0.202
σ_ϵ	0.790	0.807	-0.017	0.77	0.783	-0.014
c	0.705	0.719	-0.027	-	-	-

* Diff marks the difference between the non-rounded values of the parameter estimates for the HUNT2 cohort and the estimates on the HUNT3 cohort.

4.4 Validation of Model Predictions for the HUNT3 Cohort

The validation is performed as described in Section 3.7 on the HUNT3 cohort. However, due to privacy, we present individual results on simulated data given in Table 4.2. The posterior predictive distributions for the three simulated participants are plotted in Figure 4.6 and Figure 4.7. From this simple example, it seems the posterior predictive distribution of the SPM is shifted towards larger values than the naive model, and more so for Id1 and Id3, who have more extreme explanatory variables. When predicting the probability of dropping out, the SPM has a much wider distribution than the naive model. This difference, however, can be explained by the random effect ϵ_i, which is not present in the naive model. The posterior means of the probability of dropping out are similar when comparing the SPM with the naive model.

On a population basis, we compare the CRPS scores of the posterior predictive distributions of BP_{Fi} and the Brier score for the posterior predictive mean of the probability of dropping out for the HUNT3 cohort. The CRPS scores can be seen in Table 4.4 and are almost equal. The naive model performs slightly better. This can be explained by the fact that the missing participants do not affect the likelihood of the naive model. Hence this model is optimized to perform well for the present participants. The Brier scores can be seen in Table 4.5 and are very similar when we look at all participants

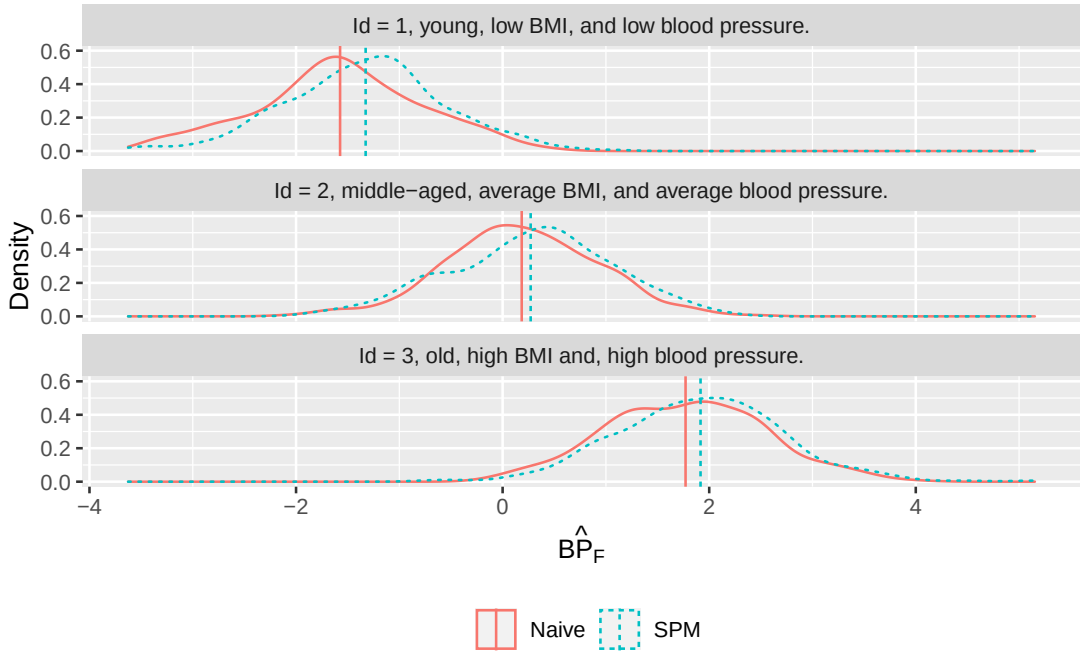


Figure 4.6.: Posterior predictive distribution of BP_F for three simulated participants with values for BP_I, age, and BMI as given in Table 4.2. The vertical lines indicate the posterior means.

Table 4.4.: CRPS score for predictions of systolic blood pressure in HUNT4 for the present participants based on the HUNT3 cohort. Bold marks the best score.

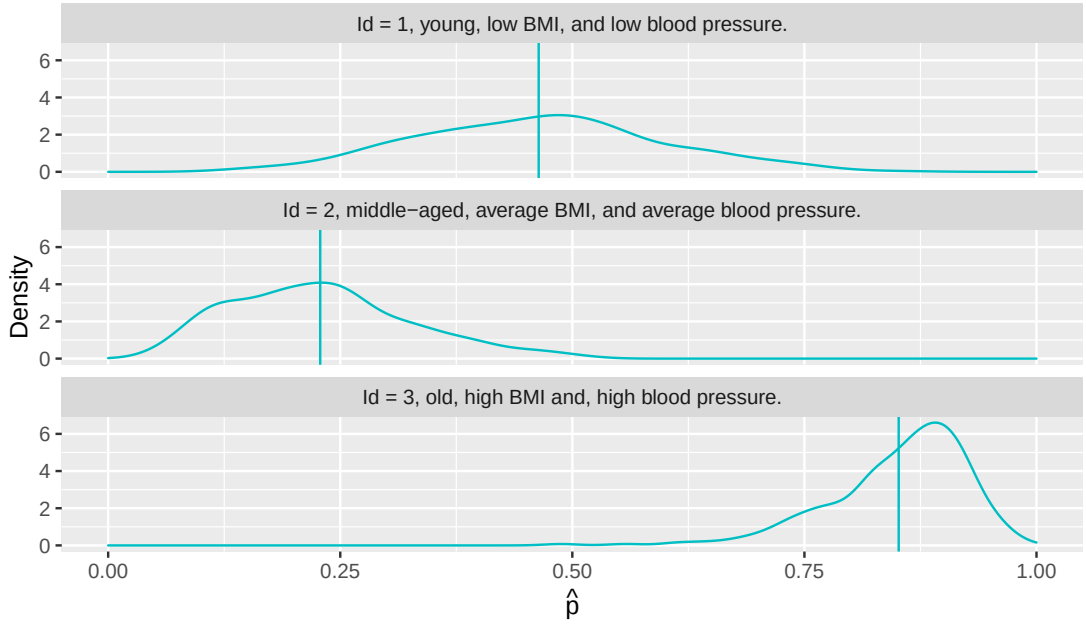
	SPM	**Naive**
CRPS	0.4406	**0.4337**

together. However, when grouped by missing status, the naive model performs better on the present participants, and the SPM performs better on the dropouts.

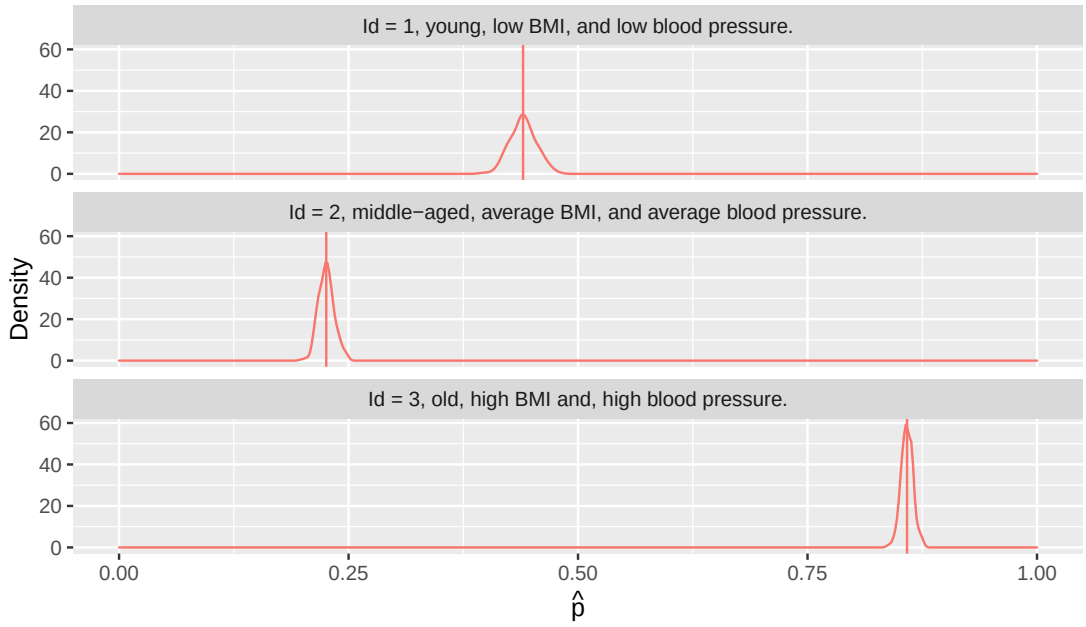
The distribution of posterior means for all participants both for the blood pressure model and the dropout model can be seen in Figure 4.8 and Figure 4.9. We see that the SPM predicts slightly higher values for the BP_F than the naive model. For the present participants, the observed distribution of BP_F is wider than the distribution of mean predictions. This distribution is only added for reference and is expected to be wider compared to a distribution of posterior mean estimates. The distributions of mean predictions of the probability of dropping out are also similar. The SPM predicts a

Table 4.5.: Brier score for mean predictions of the probability of drop out in HUNT4. All predictions are based on the HUNT3 cohort. Bold marks the best score.

	Brier		
	All	Present	Missing
SPM	0.2082	0.1656	**0.2937**
Naive	**0.2072**	**0.1602**	0.3014



(a) SPM



(b) Naive model

Figure 4.7.: Posterior predictive distribution of the probability of dropping out, p, for three simulated participants with values for BP_I, age, and BMI as given in Table 4.2.

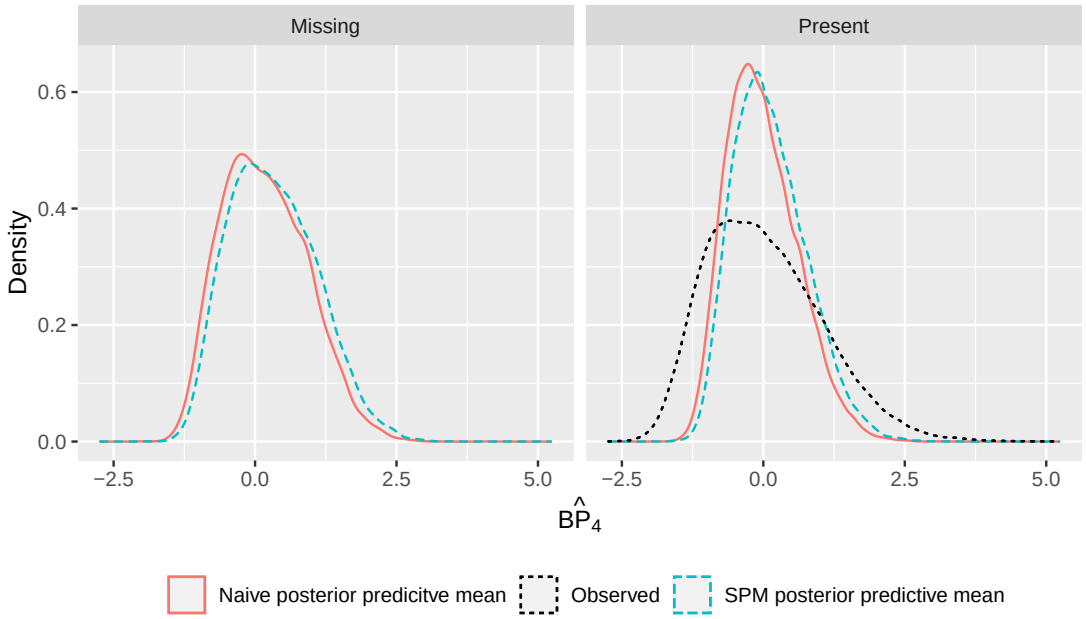


Figure 4.8.: Density of the posterior mean of BP_{Fi} for all participants grouped by missing status in the HUNT3 cohort.

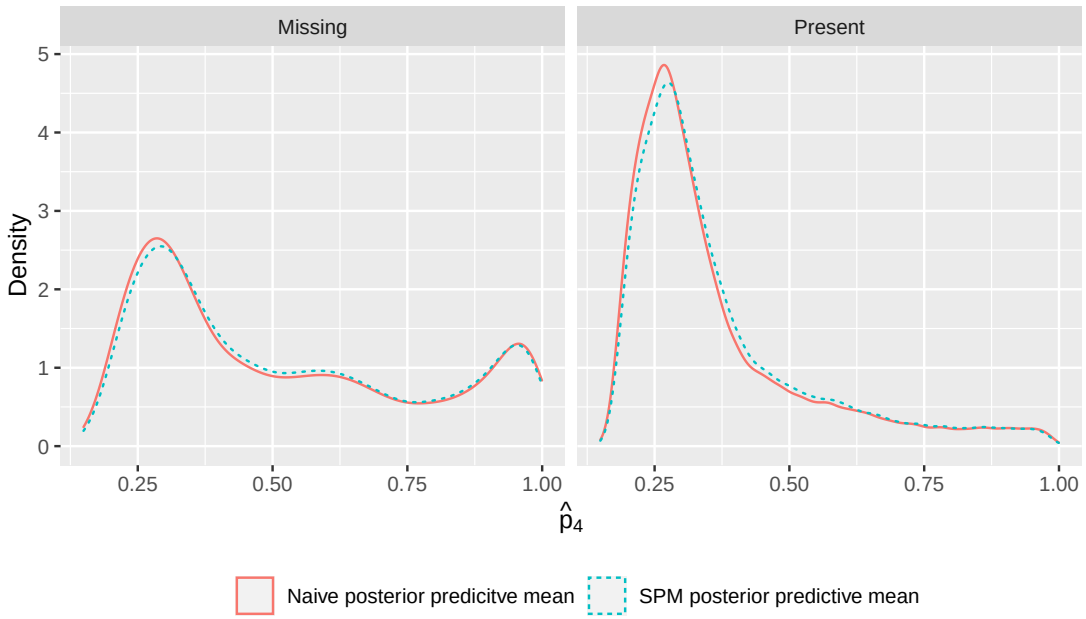


Figure 4.9.: Density of the posterior mean of p_i for all participants grouped by missing status in the HUNT3 cohort.

slightly higher probability of dropping out, on average, than the naive model. We not that the differences have little practical implication on a populational level while there are some differences on an individual level.

4.5 Validation of the Missing Not at Random Assumption

After refitting the SPM, as described in Section 3.8 both given the missing status and not, we obtain the mean absolute error (MAE) of the posterior mean predictions of each participant. The results are given in Table 4.6. We see the errors are similar. The predictions for $\boldsymbol{BP_4}$ given $\boldsymbol{m_4}$ yield a slightly smaller error than not knowing $\boldsymbol{m_4}$, suggesting some information from the missing process affects the BP_F.

Table 4.6.: Mean absolute error for $\hat{\boldsymbol{BP}}_{\boldsymbol{F}}|m$ and $\hat{\boldsymbol{BP}}_{\boldsymbol{F}}$. The best score is indicated in bold.

	$\hat{\boldsymbol{BP}}_{\boldsymbol{F}}\|m$	$\hat{\boldsymbol{BP}}_{\boldsymbol{F}}$	MAE($\hat{\boldsymbol{BP}}_{\boldsymbol{F}}\|m$) - MAE($\hat{\boldsymbol{BP}}_{\boldsymbol{F}}$)
MAE	**0.6140**	0.6155	-0.0014

Simulation Studies

5

We set up several simulation studies to explore the properties of the SPM, the naive model and the method validating MNAR by conditioning on missing status for datasets with the size and structure of the HUNT3 cohort. All simulation studies regarding bias and coverage of the models are preformed with the same number of participants and explanatory variables X as in the HUNT2 cohort, i.e. $X = X_{HUNT2}$. Further, when simulating data, the parameters $\boldsymbol{\theta} = \boldsymbol{\theta_{true}}$ are set to the posterior mean estimates of the SPM and naive model fitted on the HUNT2 cohort, see Table 4.1. When predictions are studied, simulated validation datasets are based on the same size and explanatory variables as the HUNT3 cohort, i.e. $X = X_{HUNT3}$ is used.

5.1 Simulation Studies Exploring Bias and Coverage

We performed several simulation studies to evaluate how well the SPM and naive model estimates known parameters on MNAR and MAR data. We do this to evaluate if the models are biased or able to reproduce the original parameter values.

5.1.1 Setup

We use the parameter estimates

$$\boldsymbol{\theta}_{true} = \pi(\hat{\alpha}_0, \hat{\alpha}_{BP}, \hat{\alpha}_{age}, \hat{\alpha}_{BMI}, \hat{\alpha}_{sex}, \hat{\beta}_0, \hat{\beta}_{BP}, \hat{\beta}_{BMI}, \hat{\beta}_{sex}, \hat{f}(.), \hat{\sigma}_\epsilon), \tag{5.1}$$

obtained from both the main SPM (3.3) and the naive model (3.4) displayed in Section 4.2 (Table 4.1) and the original explanatory variables X_{HUNT2} from the HUNT2

Table 5.1.: This table indicates which parameter estimates and data missing mechanism are used, in each study, to construct the new values for $\boldsymbol{BP_F}$ and $\boldsymbol{m}$. SPM signifies the we use the posterior means for all parameters from the SPM while naive signifies that we use the posterior means obtained by the naive model (Table 4.1). We also display the true value of the association parameter in the respective cases.

Study	**Parameter estimates**	**Missing mechanism**	$\hat{c}$
1	SPM	MNAR	0.705
2	SPM	MAR	0
3	Naive	MNAR	0.705
4	Naive	MAR	0

cohort to simulate new values for $Y = (\boldsymbol{BP_F}, \boldsymbol{m})$, denoted Y^{new}. This gives us new simulated data sets (X_{HUNT2}, Y_{new}) which are either MNAR or MAR depending on the value of $\hat{c}$. When $\hat{c} = 0.705$ the data are MNAR and when $\hat{c} = 0$ data is MAR.

In total, we perform four simulations studies exploring model bias and coverage with a combination of parameter estimates and missing mechanisms as given in Table 5.1. All four studies follow the same setup given as follows:

Run 100 simulations where each simulation is indicated by index l.

For $l = 1 : 100$:

– Create new values for Y denoted Y_l for all participant based on the original explanatory variables, X_{HUNT2}, from the HUNT2 cohort, and parameter estimates $\boldsymbol{\theta}_{true}$ as follows:

$$\begin{aligned} \epsilon_{il} &\sim N(0, \hat{\sigma}_\epsilon) \\ BP_{Fil} &= \hat{\alpha}_0 + \hat{\alpha}_{BP} BP_{2i} + \hat{\alpha}_{age} age_{2i} + \hat{\alpha}_{BMI} BMI_{2i} + \hat{\alpha}_{sex} sex_{2i} + \epsilon_{il} \\ \text{logit}(p_{il}) &= \hat{\beta}_0 + \hat{\beta}_{BP} BP_{2i} + \hat{f}(age_{2i}) + \hat{\beta}_{BMI} BMI_{2i} + \hat{\beta}_{sex} sex_{2i} + \hat{c}\epsilon_{il} \\ m_{il} &\sim \text{Bernoulli}(p_i). \end{aligned} \tag{5.2}$$

We recall that the index 2 signifies that the data is from the HUNT2 study. This gives us a new simulated dataset $\text{data}_{new} = (X_{HUNT2}, Y_l)$ where $Y_l = (\boldsymbol{BP_{F_l}}, \boldsymbol{m_l})$.

– Fit the SPM to data_{new} and summarize the simulation by the posterior mean of all k model parameters, θ_{kl}, and the binary variable $cover_{kl}$ given by:

$$\text{cover}_{kl} = \begin{cases} 1 & \text{if } \hat{\theta}_{k_{true}} \text{ falls within the } 95\% \text{ equal tailed credible interval} \\ & \text{of the posterior distribution of } \theta_{kl} \\ 0 & \text{otherwise.} \end{cases} \tag{5.3}$$

When we have 100 estimates and cover values for all model parameters we compute the mean of the posterior means

$$\bar{\theta_k} = \frac{1}{100} \sum_{l=1}^{100} \theta_{kl}$$

for all k parameters and compute the bias of the posterior means,

$$\text{Bias } (\bar{\theta_k}) = \frac{\sum_{l=1}^{100} (\theta_{kl} - \hat{\theta}_{k_{true}})}{100}.$$

Finally we compute the coverage defined as

$$\text{coverage}_k = \frac{1}{100}\sum_{l=1}^{100}\text{cover}_{kl},$$

where $cover_{kl}$ is defined in (5.3).

5.1.2 Results and Discussion

The resulting mean posterior mean, bias, and coverage for all simulation studies can be viewed in Table A.1, Table A.2, Table A.3, and Table A.4 in Appendix A.

The distribution of the posterior means from the simulation studies can be seen in Figure 5.1, Figure 5.2, Figure 5.3, and Figure 5.4. We see from Figure 5.1 and Figure 5.3 that both the naive model and SPM are biased when the data is MNAR. However, the SPM is less biased than the naive model especially for the blood pressure model (i.e. $\alpha_0, \alpha_{PB}, \alpha_{age}, \alpha_{sex}$, and α_{BMI}). Still, there are some reproducibility issues. These are especially pronounced for the association parameter c, which has low coverage (Table A.1 and Table A.3).

When the data is MAR, both the SPM and naive model have very little bias and give similar results for most parameters, as seen in Figure 5.2 and Figure 5.4. Also, the association parameter c is centered around zero and has good coverage. This finding reassures our assumption that the SPM and the naive model perform similarly if the data are MAR, although the naive model is slightly better.

When comparing all four studies, we see that both models are less biased when the data is MAR than when it is MNAR. Missing at random data also results in better coverage for most parameters than when the data is MNAR(Table A.1, Table A.2, Table A.3, and Table A.4). If one does not know if the data is MAR or MNAR, the naive model produces a larger error when fitted to MNAR data than the SPM produces when fitted to data MAR. It seems the SPM is more robust at modeling data both MAR and MNAR while the naive model performs poorly on data MNAR. This finding supports using a SPM when the data is likely to be MNAR.

To conclude, both models are biased when the data is MNAR. However, the SPM is less biased and has much better coverage for most parameters (Table A.1 and Table A.3). The two models perform almost identical regarding bias and coverage when the data is MAR (Table A.2 and Table A.4), although the naive model is slightly better in terms of a lower bias.

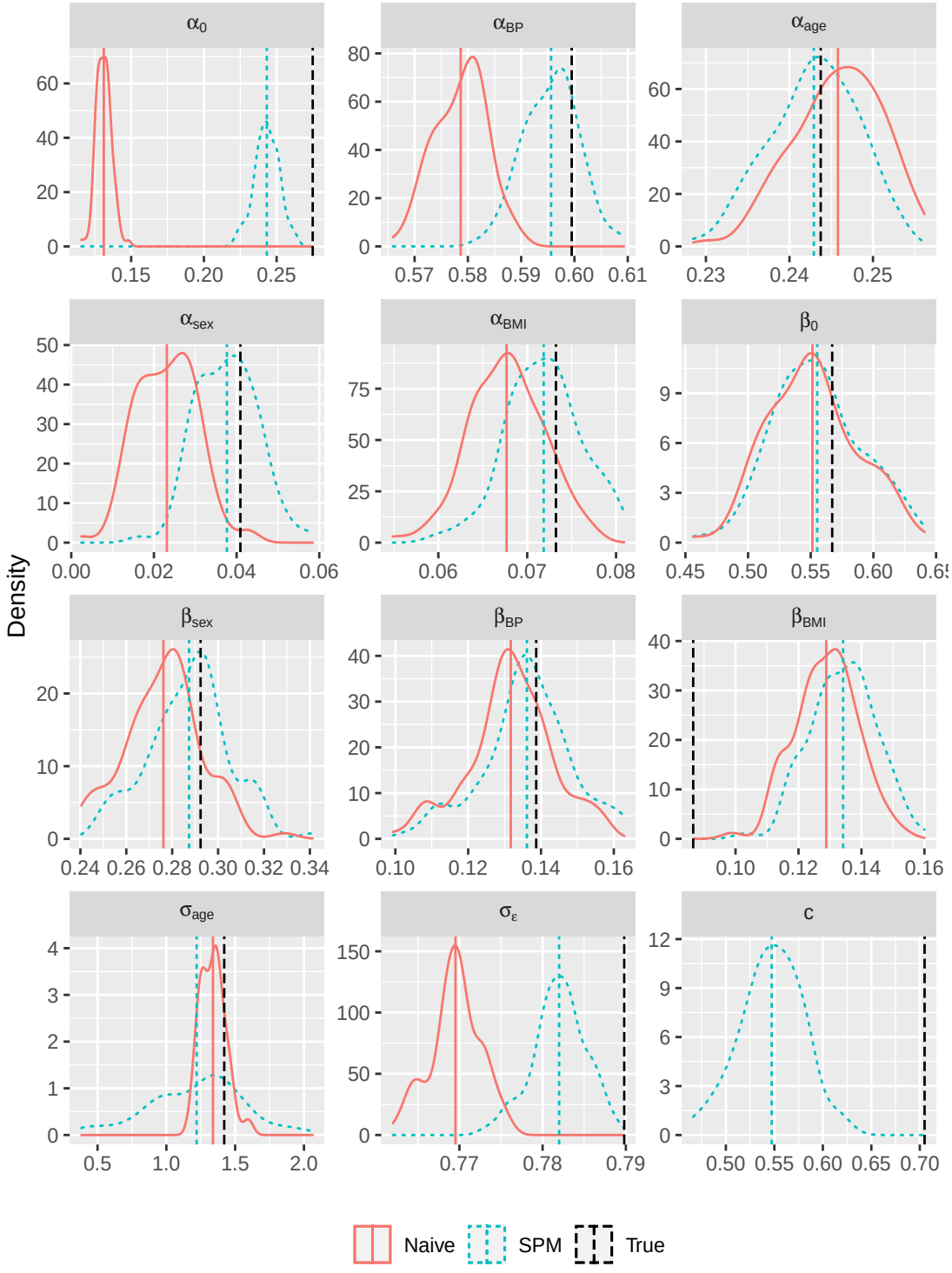


Figure 5.1.: Resulting distribution of posterior mean estimates after 100 iterations of model fits on simulated MNAR data following the SPM (Study 1, Table 5.1) as described in Section 5.1.1. The vertical dotted black line indicates the true values for all model parameters. Furthermore, the mean of posterior means for both the SPM and naive model are indicated by red and blue vertical lines, respectively.

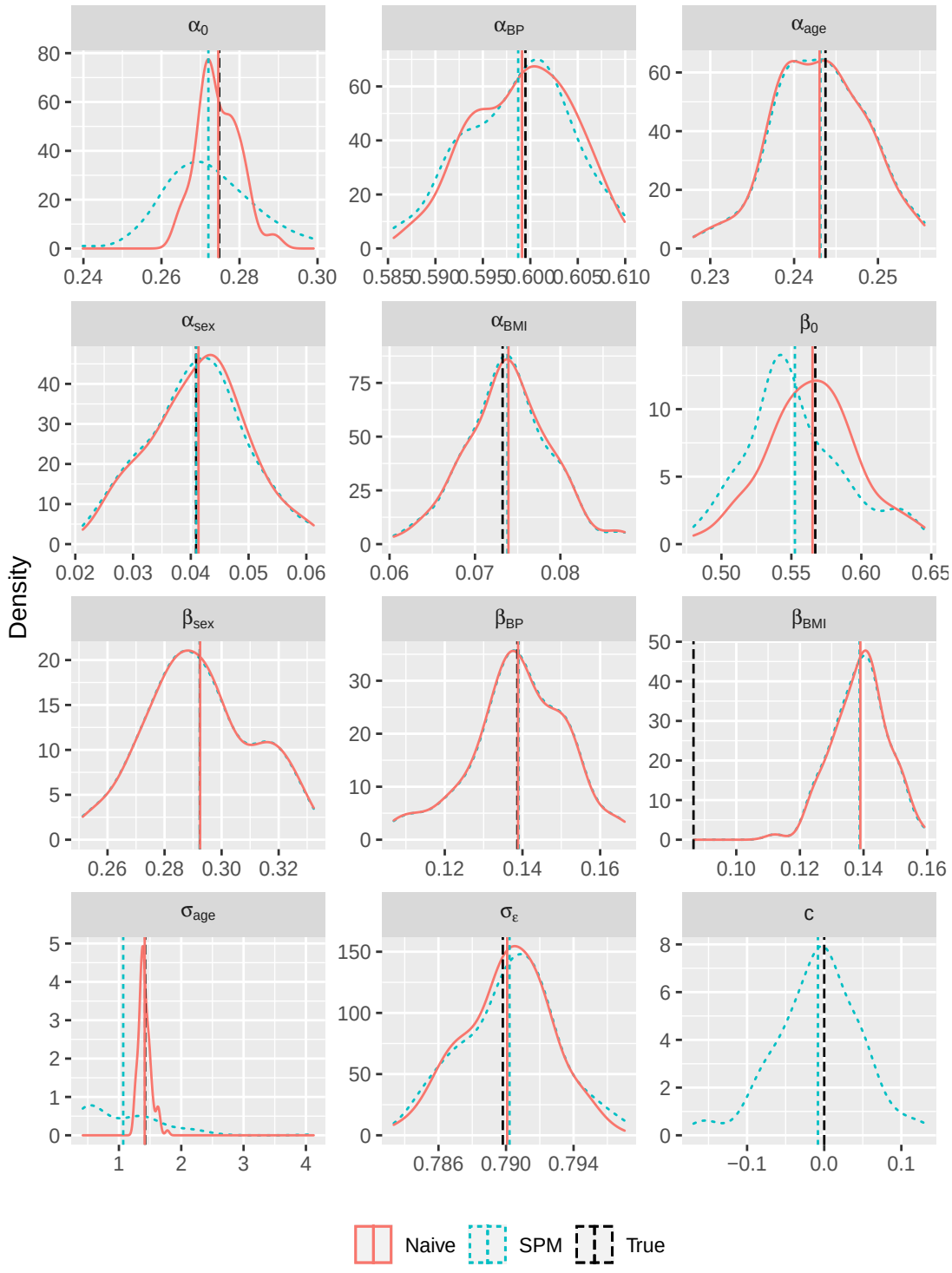


Figure 5.2.: Distribution of posterior mean estimates after 100 iterations of model fits on simulated MAR data following the SPM (Study 2, Table 5.1) as described in Section 5.1.1. The vertical dotted black line indicates the true values for all model parameters. Furthermore, the mean of posterior means for both the SPM and naive model are indicated by red and blue vertical lines, respectively.

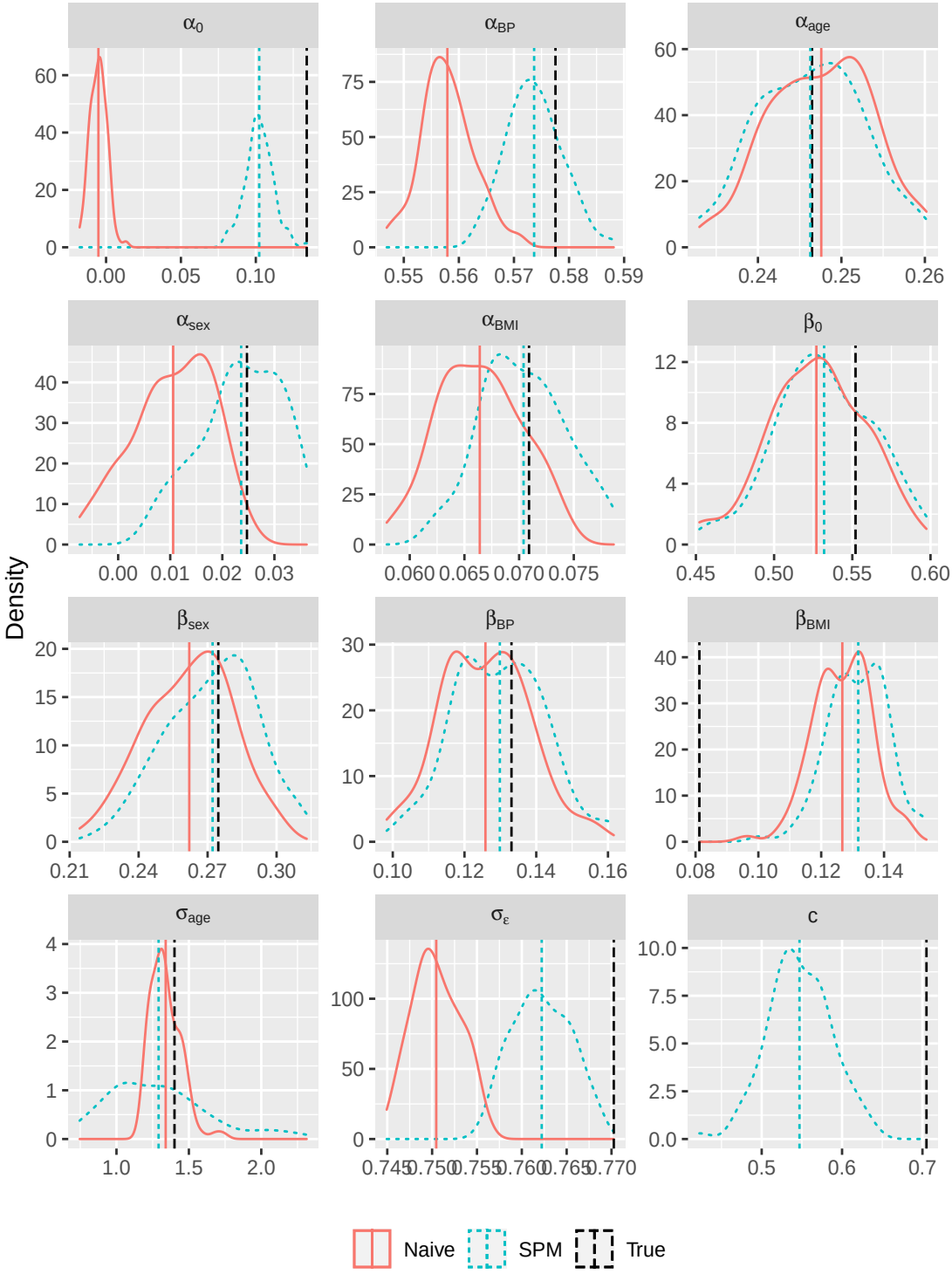


Figure 5.3.: Resulting distribution of posterior mean estimates after 100 iterations of model fits on simulated MNAR data following the naive model (Study 3, Table 5.1) as described in Section 5.1.1. The vertical dotted black line indicates the true values for all model parameters. Furthermore, the mean of posterior means for both the SPM and naive model are indicated by red and blue vertical lines, respectively.

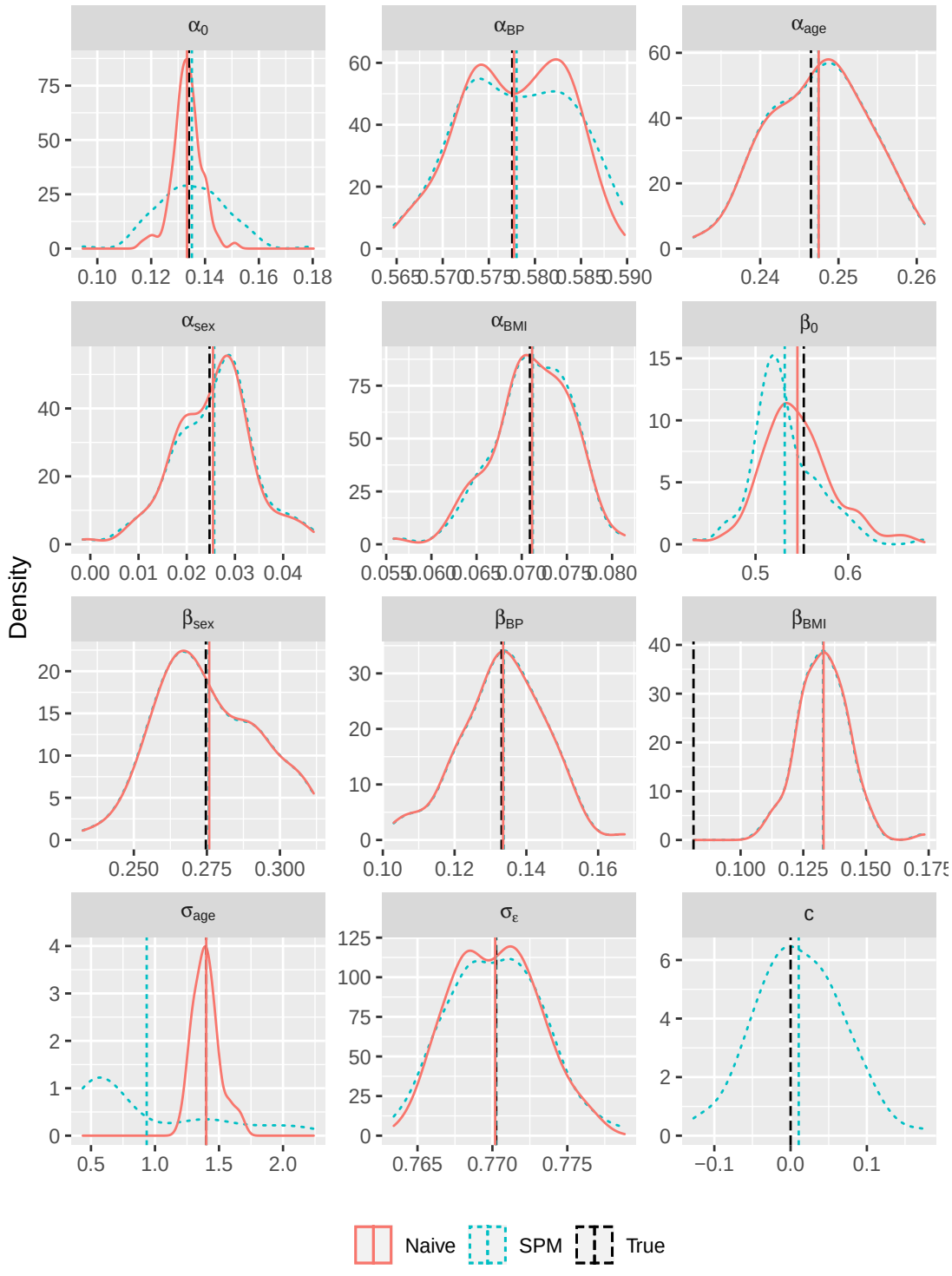


Figure 5.4.: Distribution of posterior mean estimates after 100 iterations of model fits on simulated MNAR data following the naive model (Study 4, Table 5.1) as described in Section 5.1.1. The vertical dotted black line indicates the true values for all model parameters. Furthermore, the mean of posterior means for both the SPM and naive model are indicated by red and blue vertical lines, respectively.

5.2 Validation of Model Predictions

To test the validation scheme introduced in Section 3.7 comparing the CRPS and Brier score of the model predictions for the SPM and naive model, we perform a simulation study with known data. The simulated validation datasets use explanatory variables from the HUNT3 cohort, i.e. $X = X_{HUNT3}$, and simulate new values for Y based on the posterior mean estimates of the SPM. A large advantage for the simulated data sets is that predictive performance can be evaluated not only for present participants, but also for those that are missing, which we include.

5.2.1 Setup

We use the posterior mean estimates $\boldsymbol{\theta}_{true}$ obtained by the SPM (Table 4.1) as the true values for the model parameters analogous to Section 5.1.1 to simulate values for Y, Y_{new}. Then we refit the SPM (3.3) and naive model (3.4) to this dataset (X_{HUNT2}, Y_{new}) and denote these model fits SPM^* and $Naive^*$ model, where $*$ denotes the new model fits.

The setup of the 100 simulations is as follows:

– First we construct new values for Y similarly as in Section 5.1.1 but with explanatory variables from the HUNT3 cohort for all i participants,

$$\begin{aligned} \epsilon_{il} &\sim N(0, \hat{\sigma}_\epsilon) \\ BP_{Fil} &= \hat{\alpha}_0 + \hat{\alpha}_{BP} BP_{3i} + \hat{\alpha}_{age} age_{3i} + \hat{\alpha}_{BMI} BMI_{3i} + \hat{\alpha}_{sex} sex_{3i} + \epsilon_{il} \\ \text{logit}(p_{il}) &= \hat{\beta}_0 + \hat{\beta}_{BP} BP_{3i} + \hat{f}(age_{3i}) + \hat{\beta}_{BMI} BMI_{3i} + \hat{\beta}_{sex} sex_{3i} + \hat{c}\epsilon_{il} \\ m_{il} &\sim \text{Bernoulli}(p_{il}). \end{aligned} \tag{5.4}$$

This gives us a new dataset, $\text{data}_{new} = (X_{HUNT3}, Y_l)$, with data MNAR.

– Run the validation procedure as described in Section 3.7 on data_{new} with SPM^* and $Naive*$ as model fits to obtain the mean CRPS score, $CRPS_l$, and the Brier score, $Brier_l$.

We now have 100 values of the mean CRPS and Brier score for all participants (present and missing), the present, and the missing participants for each simulated dataset for the SPM^* and $naive^*$ model.

5.2.2 Results and Discussion

Figure 5.5 displays the distribution of the difference between the mean CRPS for the SPM and the naive model. This difference is displayed for all participants (missing/present) and grouped on missing status. We see that the SPM performs better for all participants

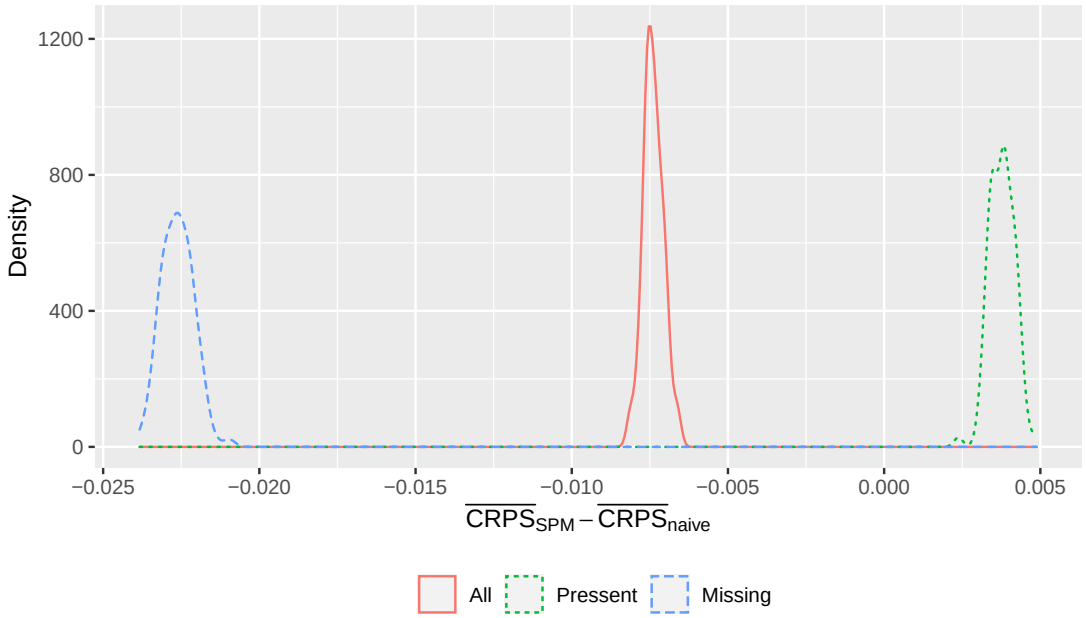


Figure 5.5.: Difference in mean CRPS score for the $SPM*$ and $naive*$ model on 100 simulated dataset following the scheme described in Section 5.2.1.

and the dropouts, but the naive model performs better on present participants. This confirms our previous suspicion that even if the data is MNAR and the SPM fits the actual data better, the naive model is expected to obtain a better CRPS score when only tested on the present participants.

Figure 5.6 shows the distribution of the difference in Brier score for all participants, the dropouts, and the present participants. We see similar results here, although the Brier score for all participants, in this case, covers zero. However, for the dropout process, the original parameter estimates of the SPM and naive model (Table 4.1) were much more similar than for the blood pressure process. Naturally, these predictions will then be closer to one another.

To summarize, we find that the CRPS score for the blood pressure process can be better for the naive model than the SPM if we only consider the present participants. Otherwise, the CRPS of the SPM is the best if the data is MNAR and follows the SPM. The Brier score of the dropout process for the SPM is better for dropouts, very similar for all participants (present and missing), and worse for the present participants than the naive model on MNAR data following the SPM.

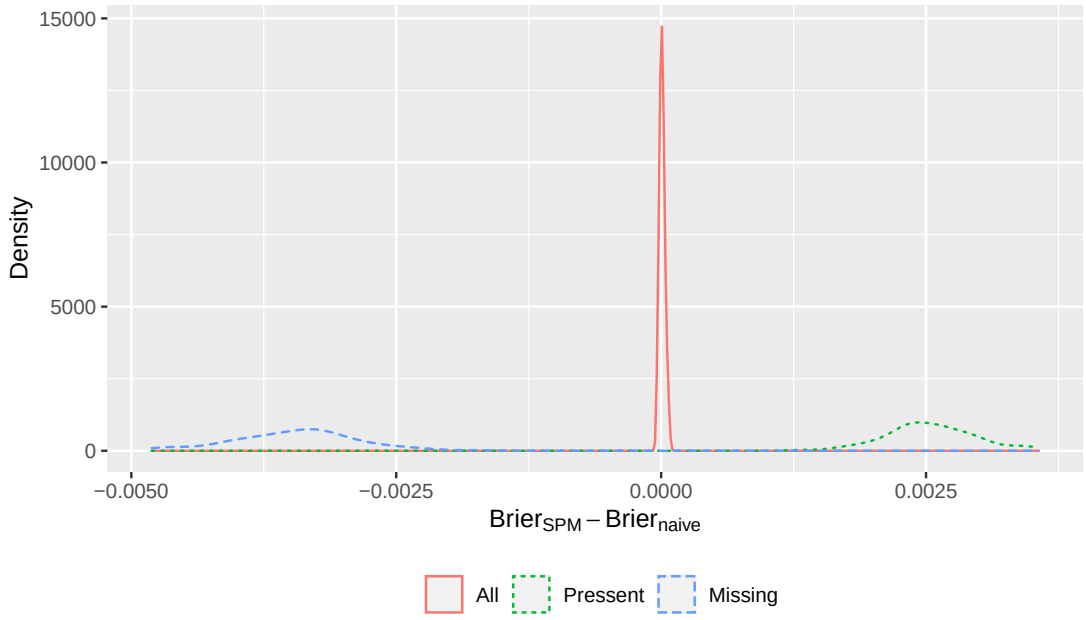


Figure 5.6.: Difference in Brier score for the $SPM*$ and $naive*$ model on 100 simulated dataset following the scheme described in Section 5.2.1.

5.3 Validation of Evaluation Method of Data Missing Not at Random

In section Section 3.8 we proposed a novel validation scheme for assessing if the data is MNAR when it follows the SPM. We suspect that data MNAR means the missing status contains information about $\boldsymbol{BP_F}$. This suspicion would imply predicting $\hat{BP}_{Fi}|m_i$ should result in better predictions than $\hat{BP}_{Fi}$.

When we implement the scheme introduced in Section 3.8, we have no control of the underlying data missing mechanism of the real data. Therefore, we simulate the data to be both MNAR and MAR to evaluate how much difference we can expect to obtain between the mean absolute errors (MAE) of $\boldsymbol{\hat{BP}_F}|\boldsymbol{m}$ and $\boldsymbol{\hat{BP}_F}$ when the data is MNAR and MAR respectively. For simplicity, we use the posterior mean prediction of $\hat{BP}_{Fi}|m_i$ and $\hat{BP}_{Fi}$.

5.3.1 Setup

We describe the setup when the data is simulated MNAR and MAR separate to avoid confusion on how the data is created.

Data simulated Missing Not at Random

We start by using the posterior mean estimates obtained by the SPM (3.3) displayed in Table 4.1 and the original explanatory variables from the HUNT2 cohort to simu-

late new values for Y, denoted Y_{new}, as done in Section 5.1.1 and fit a new SPM to (X_{HUNT2}, Y_{new}) as done in Section 5.2.1. We call this model fit SPM_{MNAR} since it is fitted to data MNAR.

Further, we run 100 iterations, indicated by index $l = 1 : 100$, where we do the following:

– Simulate a new dataset as done in (5.4) by using explanatory variables from the HUNT3 cohort and posterior mean estimates from the model parameters of the main SPM (3.3) presented in Table 4.1 to create Y_l as done in (5.4).

– Use SPM_{MNAR} to predict $B\hat{P}_{Fil}|m_{il}$ and $B\hat{P}_{Fil}$ for all i participants as described in Section 3.8

– Compute $MAE(\boldsymbol{B\hat{P}_{F_l}}|\boldsymbol{m_l})$ and $MAE(\boldsymbol{B\hat{P}_{F_l}})$

Data simulated Missing at Random

We repeat the what we did for data MNAR for data MAR by using the posterior mean estimates θ_{true} obtained by the naive model (3.4) (Table 4.1) with explanatory variables from the HUNT2 cohort as done in Section 5.1.1 to create Y_{new}. We fit a SPM to (X_{HUNT2}, Y_{new}) and call it SPM_{MAR} since it is fitted to data MAR following the naive model.

Then we run this simulation procedure 100 times, indicated by index $l = 1 : 100$, where we do the following:

– Create a new dataset by using explanatory variables from the HUNT3 cohort and posterior mean estimates from the model parameters from the original naive model presented in Table 4.1 to create Y_l.

– Use SPM_{MAR} to predict $B\hat{P}_{F_{il}}{}'|m_{il}$ and $B\hat{P}_{F_{il}}{}'$ as described in Section 3.8. We use a $'$ to indicate that the data used is MAR.

– Compute $MAE(\boldsymbol{B\hat{P}_{F_l}}{}'|\boldsymbol{m_l})$ and $MAE(\boldsymbol{B\hat{P}_{F_l}}{}')$.

5.3.2 Results and Discussion

Data Simulated Missing Not at Random

We can see from Figure 5.7 that the distribution of MAEs for the predictions of Y_l in the 100 simulated datasets is shifted towards lower values when the missing status is known when the data is MNAR. Further, we see from Figure 5.9 that the mean absolute error was smaller for every simulated dataset when predicting $\boldsymbol{B\hat{P}_F}|\boldsymbol{m}$ than $\boldsymbol{B\hat{P}_F}$ since zero is not contained in the distribution of $MAE(\boldsymbol{B\hat{P}_F}|\boldsymbol{m} - \boldsymbol{B\hat{P}_F})$. Hence, for data MNAR following the SPM, we can expect the predictions of $\boldsymbol{B\hat{P}_F}|\boldsymbol{m}$ to be better than $\boldsymbol{B\hat{P}_F}$.

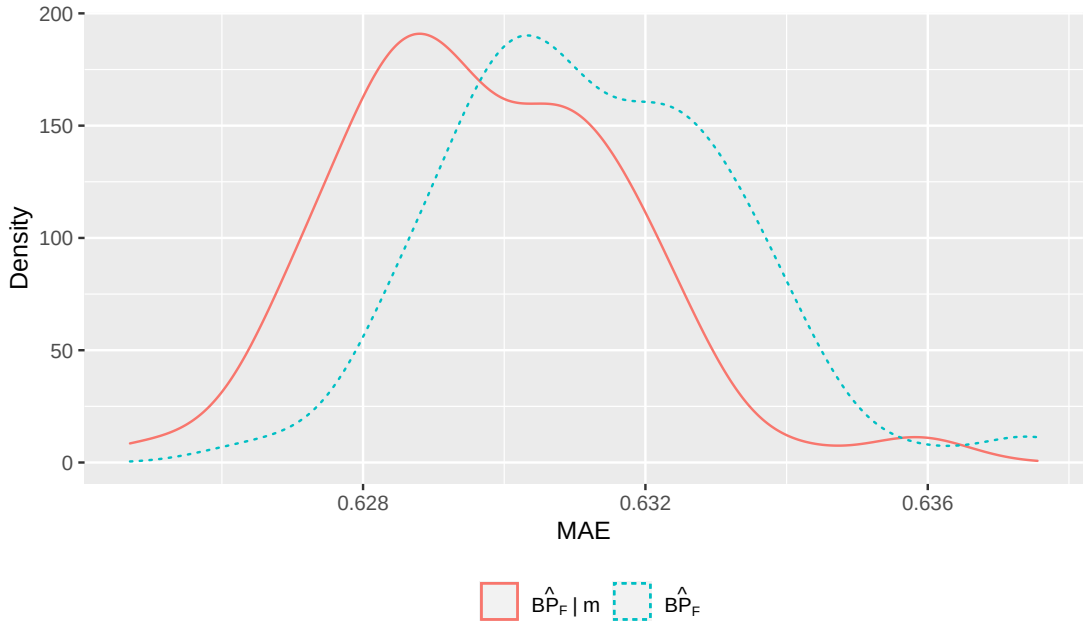


Figure 5.7.: Distribution of the mean absolute error (MAE) of the posterior mean prediction $\boldsymbol{B\hat{P}_F}|\boldsymbol{m}$ and $\boldsymbol{B\hat{P}_F}$ for future blood pressure obtained from the validation scheme presented in Section 5.3.1 when the data is MNAR.

Data Simulated Missing at Random

Figure 5.8 shows that when the data is MAR, we obtain almost the same mean absolute error for for the predictions of Y_l in every simulated dataset when predicting $\boldsymbol{B\hat{P}_F}'|\boldsymbol{m}$ as when predicting $\boldsymbol{B\hat{P}_F}'$. In addition we see from Figure 5.9 that the distribution of $MAE(\boldsymbol{B\hat{P}_F}'|\boldsymbol{m}) - MAE(\boldsymbol{B\hat{P}_F}')$ covers zero.

Comparing Data Missing Not at Random with data Missing at Random

From Figure 5.9 we also see that there is no overlap between the two distributions, $MAE(\boldsymbol{B\hat{P}_F}|\boldsymbol{m}) - MAE(\boldsymbol{B\hat{P}_F})$ and $MAE(\boldsymbol{B\hat{P}_F}'|\boldsymbol{m}) - MAE(\boldsymbol{B\hat{P}_F}')$. The distributions are also very narrow both when the data is MNAR and MAR. This demonstrates that even minor differences in MAE between predictions of $\boldsymbol{B\hat{P}_F}|\boldsymbol{m}$ and $\boldsymbol{B\hat{P}_F}$ indicate that data are MNAR.

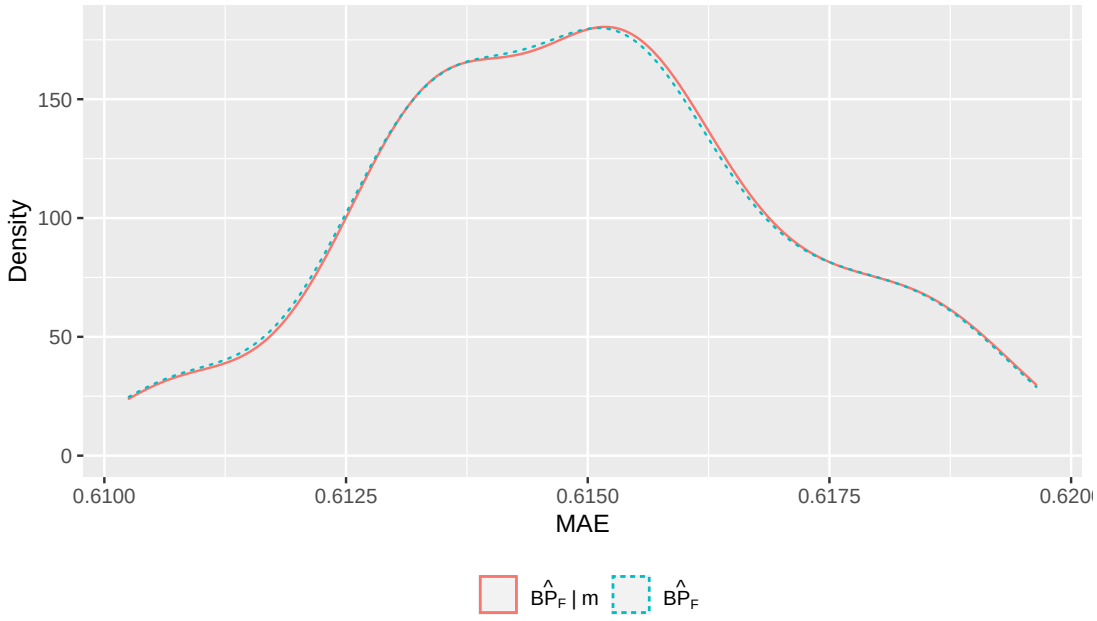


Figure 5.8.: Distribution of the mean absolute error (MAE) of the posterior mean prediction $\boldsymbol{B\hat{P}_F}'|\boldsymbol{m}$ and $\boldsymbol{B\hat{P}_F'}$ for future blood pressure obtained from the validation scheme presented in Section 5.3.1 when the data is MAR.

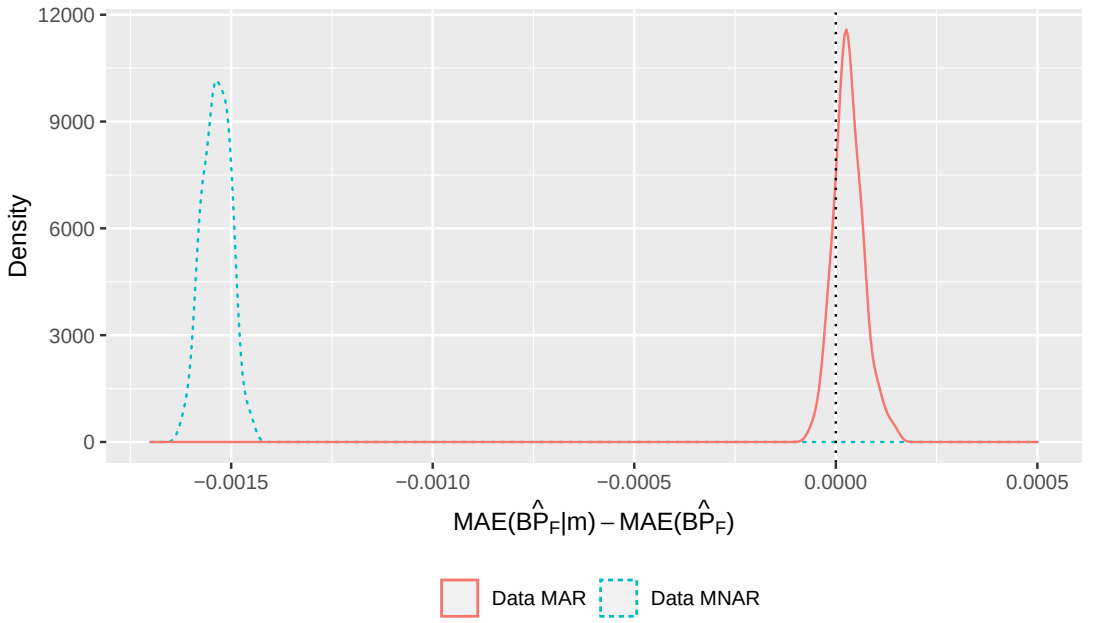


Figure 5.9.: Difference in mean absolute error (MAE) of the predictions of $\boldsymbol{BP_F}|\boldsymbol{m}$ and $\boldsymbol{BP_F}$ when the data is MNAR and MAR. The black vertical line indicates zero.

Prior Sensitivity Analysis

6

The simulation studies performed in Section 5.1 uncovered identifiability issues for both the naive model and the SPM when the data was MNAR (Figure 5.3 and Figure 5.1). Even though the SPM is less biased than the naive model, there is still a substantial bias. The coverage for the association parameter c is especially low (Table A.3 and Table A.1). The association parameter is of great importance in the SPM since it defines the connection between the dropout process and the blood pressure process. Therefore, we were left with the question; Could misspecification of the informative prior for the association parameter c be the cause of some of this bias and low coverage? This question motivated a prior sensitivity analysis. All the models fitted in Chapter 4 used a normal prior with mean zero and standard deviation one. In this analysis, we fit the main SPM defined in (3.3) with different choices for the prior for c displayed in Table 6.1. All other priors are kept identical to the specification given in Section 3.3.3. We use real data from the HUNT2 cohort.

Figure 6.1 shows the resulting 95% equi-tailed credible intervals for all parameters. We clearly see that the latent field, α_0, α_{BP}, α_{age}, α_{BMI}, α_{sex}, β_0, β_{BP}, β_{BMI}, β_{sex}, c, have almost identical credible intervals. Even when the prior for c is misspecified with mean ten and variance one, the SPM obtains almost identical posterior means as with the other prior distributions for c. The credible intervals for the hyperparameter σ_{age} vary slightly more. However, we see in Figure 6.2 that the resulting posterior mean of the age effect is practically identical for all the different priors. Hence we conclude that the model is not sensitive to different choices of c.

Table 6.1.: Priors for the association parameter c.

Model name	Prior for c
$mu = 0, sigma = 1$	N(0,1^2)
$mu = 0, sigma = 10$	N(0,10^2)
$mu = 0, sigma = 100$	N(0,100^2)
$mu = 1, sigma = 1$	N(1,1^2)
$mu = 1, sigma = 10$	N(1,10^2)
$mu = 1, sigma = 100$	N(1,100^2)
$mu = 10, sigma = 100$	N(10,100^2)
$mu = 10, sigma = 1$	N(10,1^2)

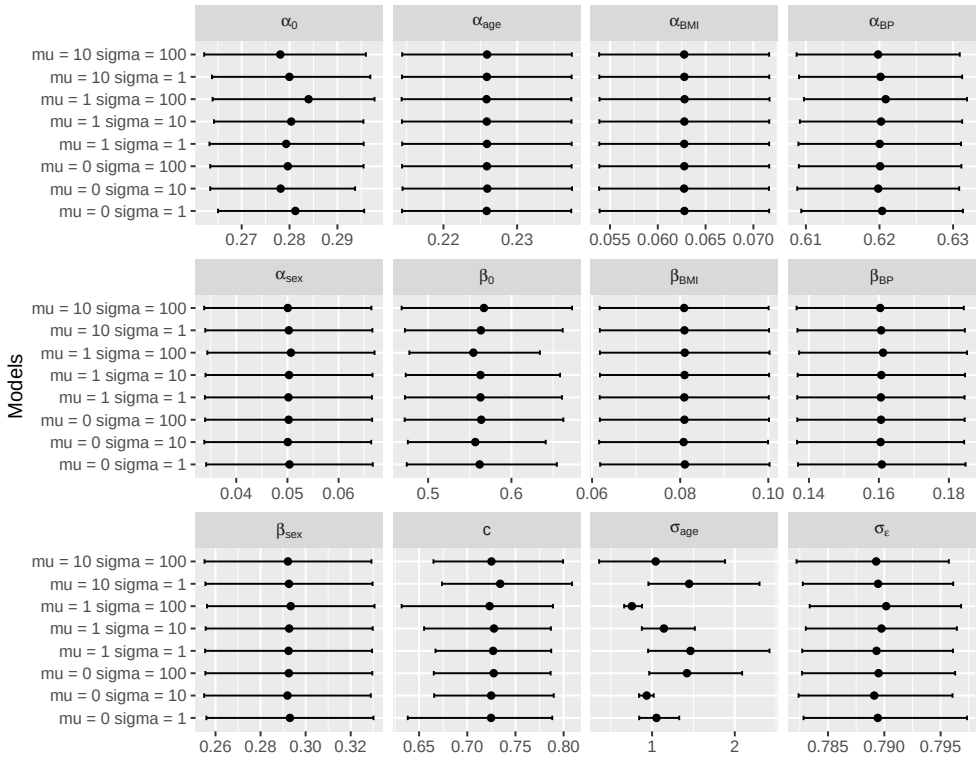


Figure 6.1.: Posterior mean and 95 % credible intervals for all model parameters of the SPM with different priors for the association parameter c. The model names are constructed so the first digit is the mean and the second the is the standard deviation used in the prior for c.

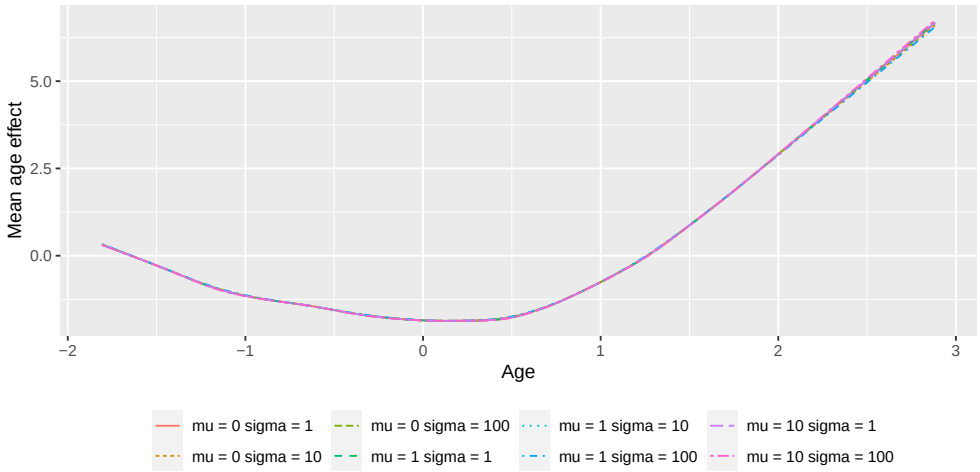


Figure 6.2.: Mean posterior age effect for all models tested in the prior sensitivity study Chapter 6

7 Simulated Data for Reproducability Purposes

To follow the standard of high-end journals such as the Journal of the American Statistical Association, we have chosen to make all code publicly available and generate a simulated dataset. These can be found in the following GitHub repository: Hofman [2021a]. For reproducibility reasons, the best would be to make all data available. However, data from the HUNT study contains personal information and can not be shared to protect the privacy of the participants. To enable third parties to evaluate the methods and findings of this work, we have constructed a simulated dataset with similar properties as the original data.

The explanatory variables are generated based on graphical inspection, summary statistics of the HUNT2 cohort, including empirical correlations, and simple linear frequentist regressions. In this chapter, we use i to index the actual participants of the HUNT2 cohort and j to index the simulated participants. The response variables, i.e., future blood pressures ($\boldsymbol{BP_F}$) and the missing status $\boldsymbol{m}$ are generated based on the posterior mean estimates of the model parameters of the SPM presented in Table 4.1. The aim has been to obtain a simulated dataset with similar properties as the HUNT2 cohort, also concerning dependence in the explanatory variables. Further formal modeling and inference regarding the models for the simulated data are outside the scope of this thesis.

7.1 Generation of the Explanatory Variables

The variable sex is drawn from a binomial distribution, and age is drawn independently from sex from a truncated mixed normal distribution. BMI is drawn from a skewed normal (SN) distribution dependent on age and sex. Further BP_I is also drawn from a SN distribution depending on sex, age, and BMI.

The following paragraph consists of a detailed explanation of the simulation process for each explanatory variable.

Sex: Sex is drawn independently for each participant j from a Bernoulli distribution with parameter p equal to the fraction of female participants in HUNT2 (Table 3.2).

$$sex_j \sim \text{Binomial}(p = 0.5303)$$

Table 7.1.: Coefficients of the linear regression $BMI_I = \alpha_0 + \alpha_{age} age_i + \alpha_{sex} sex_i + \epsilon_i$ with $\epsilon_i \sim N(0, \sigma^2)$

Variable	Mean	Std. error	Pr(>|t|)
α_0	-0.023	0.005	1 e-5
α_{age}	0.191	0.004	<2e-16
α_{sex}	0.05	0.008	1.2 e-10

Age: Age is considered independent of sex and draw from a mixed normal distribution.

$$age_j \sim 0.75N(43, 15^2) + 0.25N(75, 7^2)$$

This means, for each participant, we draw a cut parameter from a Bernoulli distribution with parameter $p = 0.75$. If the cut parameter is one, we draw age from a normal distribution with mean, $\mu = 43$, and standard deviation, $\sigma = 15$, otherwise we draw from a normal distribution with mean 75 and standard deviation 7. In addition we truncate age at $age_j = 18$ and $age_j = 105$. More precisely for every value of $age_j < 18$ we draw a new value from distribution $N(43, 15^2)$ until it was greater than 18 and for every value of $age_j > 105$ we draw a new value from $N(75, 7^2)$ distribution. Afterward, we round the age to one significant digit to replicate the format used in the HUNT Study before standardizing it. We draw age_j on the original scale (years) before standardizing it to ensure the same number of possible ages as in the HUNT2 cohort.

BMI: BMI is correlated with both age and sex in HUNT2. To be able to mimic this dependency, we first perform a linear regression on the HUNT2 cohort as follows,

$$BMI_i = \alpha_0 + \alpha_{age} age_i + \alpha_{sex} sex_i + \epsilon_i,$$
$$\epsilon_i \sim N(0, \sigma^2).$$

The result of this linear regression can be seen in Table 7.1. The distribution of BMI in the HUNT2 cohort is skewed compared to a normal distribution. Therefor we use a skewed normal (SN) distribution with density as follows;

$$f(x) = \frac{2}{\sigma}\phi(\frac{x-\mu}{\sigma})\Phi(\xi(\frac{x-\mu}{\sigma}))$$

where $\phi(.)$ denotes the standard normal density function and Φ denotes the corresponding cumulative density function.

The SN used to generate BMI_j has mean $\mu = \alpha_0 + \alpha_{age} age_j + \alpha_{sex} sex_j$, standard deviation $\sigma = 0.9$ and skewness parameter $\xi = 1.5$. The coeffcients α_0, α_{age}, and α_{sex} take the values presented in Table 7.1 from the linear regression of BMI on the HUNT2 cohort. For each simulated participant j we use the corresponding simulated age and sex.

$$BMI_j \sim SN(\mu = \alpha_0 + \alpha_{age} age_j + \alpha_{sex} sex_j, \sigma = 2, \xi = 1.5).$$

Table 7.2.: Coefficients of the linear regression $BP_I = \alpha_0 + \alpha_{sex} sex_i + \alpha_{age} age_i + \alpha_{BMI} BMI_i + \epsilon_i$ with $\epsilon_i \sim N(0, \sigma^2)$ on the HUNT2 cohort.

Variable	Estimate	Std. Error	Pr(>|t|)
α_0	-0.078	0.004	<2e-16
α_{sex}	0.167	0.006	<2e-16
α_{age}	0.516	0.003	<2e-16
α_{BMI}	0.186	0.003	<2e-16

BP_I: BP_I is correlated with age, sex, and BMI in the HUNT2 cohort. Therefore we perform a linear regression on the HUNT2 cohort as follows,

$$\begin{aligned} BP_{Ii} &= \alpha_0 + \alpha_{sex} sex_i + \alpha_{age} age_i + \alpha_{bmi} BMI_i \epsilon_i \\ \epsilon_i &\sim N(0, \sigma^2). \end{aligned}$$

The results of this regression can be seen in Table 7.2. Again the true distribution is more skewed leading us to use a SN distribution with mean $\mu = \alpha_0 + \alpha_{sex} sex_j + \alpha_{age} age_j + \alpha_{bmi} BMI_j - 0.15$, standard deviation 0.85, and skewness $\xi = 3$ and where α_0, α_{sex}, α_{age}, and $alpha_{BMI}$ are given in Table 7.2.

$$BP_I \sim SN(\mu = \alpha_0 + \alpha_{sex} sex_j + \alpha_{age} age_j + \alpha_{bmi} BMI_j - 0.15, \sigma = 0.6, \xi = 3)$$

7.2 Generating the Response Values

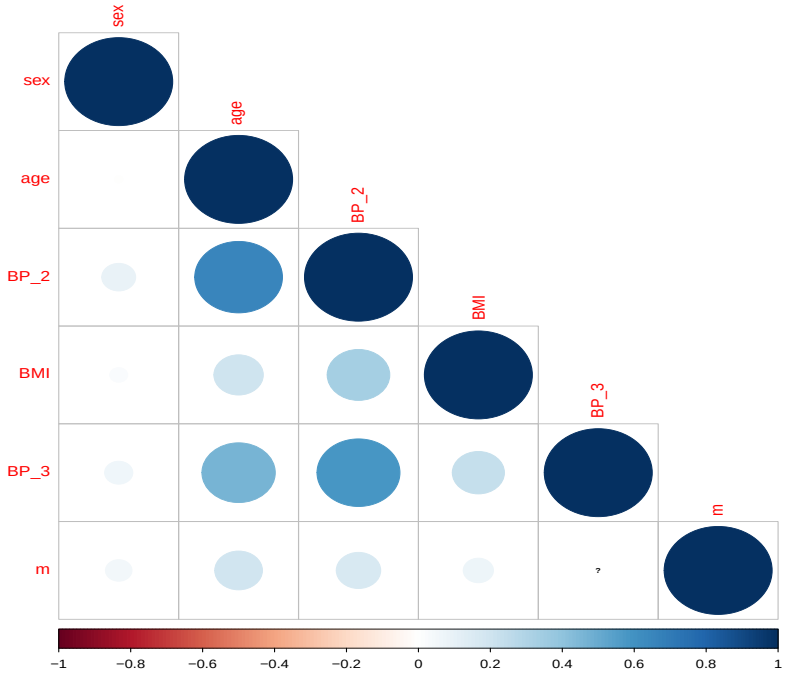
The response values BP_F and m are constructed using the posterior mean estimates for all model parameters from the SPM given in Table 4.1 fitted to the HUNT2 cohort.

$$\begin{aligned} \epsilon_j &\sim N(0, \sigma^2_{epsilon}) \\ BP_{Fj} &= \alpha_0 + \alpha_{sex} sex_j + \alpha_{age} age_j + \alpha_{bmi} BMI_j + \alpha_{BP_I} BP_{Ij} + \epsilon_j \\ \text{logit}(p_j) &= \beta_0 + \beta_{sex} sex_j + f(age_j) + \beta_{bmi} BMI_j + \beta_{BP_I} BP_{Ij} + c\epsilon_j \\ m_j &\sim \text{Bernoulli}(p_j) \end{aligned}$$

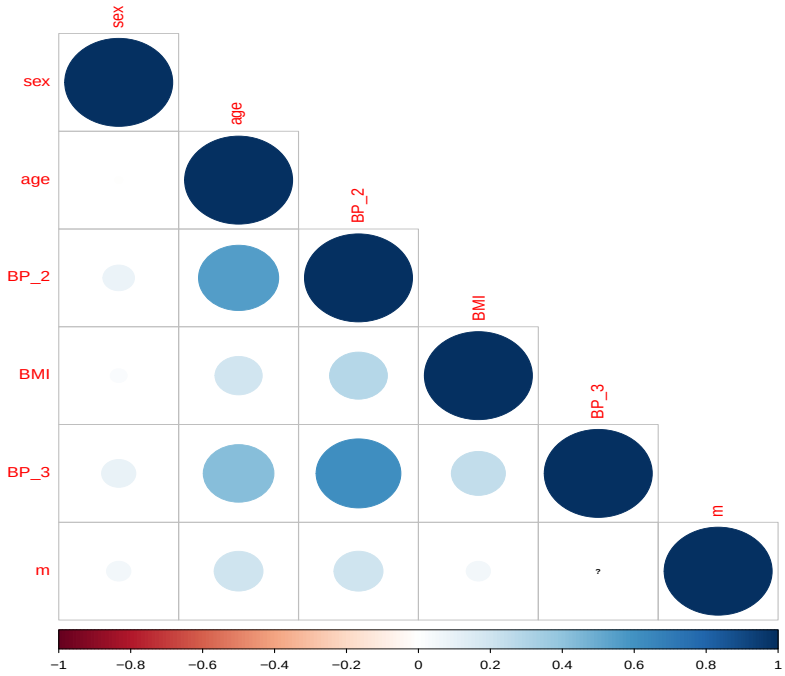
7.3 Comparison Between Simulated Data and the HUNT2 Cohort

As we can see from Figure 7.2 the distribution of age, BMI, BP_2 and BP_3 are similar. The correlation matrices for both the true and simulated data are displayed in Figure 7.1a and Figure 7.1b and also show similar properties. In Appendix B we display the posterior distribution of the latent field and the hyperparameters for the SPM (3.3) and the

naive model (3.4) for a simulated dataset of size 64385 and 1000. These datasets are available at GitHub repository Hofman [2021a]. We note that results on simulated data will not be identical to the results presented in earlier chapters. However the results for the dataset of full size, Figure B.1, Figure B.2, are similar to Figure 4.2 and Figure 4.3 presented in Section 4.2. The results for the smaller dataset, Figure B.3, Figure B.4, are not comparable as all the distributions become very wide.



(a) Correlation plot for the HUNT2 cohort.



(b) Correlation plot for simulated data.

Figure 7.1.

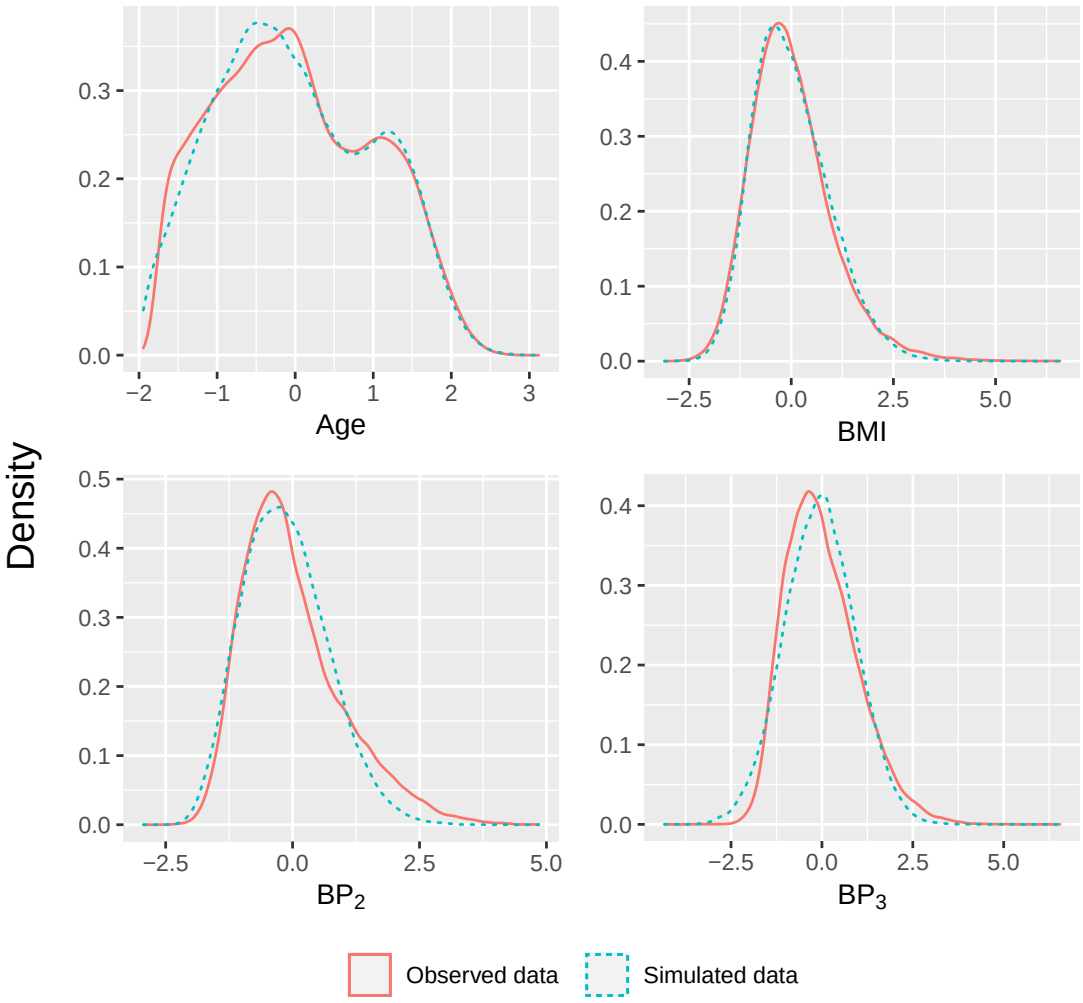


Figure 7.2.: Comparison of the distribution for age, BMI, BP_2, BP_3 from the HUNT2 cohort and simulated data.

8 Discussion

In this work, we propose reasonable models in the form of a SPM (3.3) for predicting future blood pressure (BP_F) and missing status on real and simulated data. These models are compared to a naive model (3.4) which assumes data MAR. Our findings indicate that blood pressure missing due to dropout in the HUNT Study is MNAR. The posterior distribution of the association parameter c (3.3) is non-zero, and the SPM results in different parameter estimates than the naive model (Table 4.1). The simulation studies (Section 5.1) indicate that the two models result in similar parameter estimates given MAR data. The simulation study also confirms that if the underlying assumption of data MNAR is valid, the SPM accounts for this better than the naive model, even though both models are biased. Hence, a difference in SPM and naive model parameter estimates indicate data MNAR. We note that there seem to be some identifiability issues resulting in a bias when the data is MNAR and a low coverage for some model parameters in the simulation studies (Figure 5.1,Figure 5.2, Table A.1, and Table A.2). This bias is especially pronounced for the association effect c when the data is MNAR. Our initial thought was that the low coverage could be due to misspecification of the informative prior used for c. However, the prior sensitivity study presented in Chapter 6 indicates that the model is robust to prior specifications concerning the association parameter. Hence, the source of the bias and low coverage remains unsolved. We performed all simulation studies on datasets that mimic our study systems, the HUNT2 and HUNT3 cohort, in size and explanatory variables. We have not explored the asymptotic properties of the models, such as the development of the biases with larger datasets. This is outside the scope of this thesis.

Validation of the model predictions indicated that the two models do not differ much when predicting future blood pressure and the probability of dropping out on a populational level. However, the SPM predicts both a higher $B\hat{P}_{Fi}$ and a higher probability of dropping out ($\hat{p_i}$) on an individual level. In addition, the SPM is relevant for understanding the underlying data. Through simulation studies, we see that even if the data is MNAR and the overall CRPS (for observed and missing participants) is the best for the SPM, the naive model is expected to perform better than the SPM on the present participants. Therefore we do not reject the SPM model even though the CRPS is better for the naive model than the SPM on the present participants in the HUNT3 cohort.

The new validation scheme proposed in Section 3.8 to validate the MNAR assumption, given data following the SPM, indicates that the missing status contains information about BP_F. According to the simulation studies, on this validation scheme, the model

predictions given missing status are only better than those without knowledge of missing status if the data is MNAR (Figure 5.9). When the data is MAR, the mean absolute errors of the two posterior mean predictions are almost identical. Hence we have strong indications that blood pressure missing due to dropout in the HUNT Study is MNAR.

It is worth making a note about the definition of missing. In this work, we consider every participant who did not attend the follow-up study as missing, including dead participants. One could argue that these do not have a blood pressure at the time of the follow-up study and should not be regarded as missing. However, since increased systolic blood pressure can cause death, we decided to consider these participants as missing. It could be interesting to do further research with a different definition of missingness, such as only including participants invited to the follow-up survey. One could even model two missing processes, one for the death process and one for the missing process of the still alive participants.

In this work, we have not explored the effect of measurement errors on the model results. Variables such as age and sex are not prone to measurements error. Blood pressure, on the other hand, can easily be measured incorrectly [Handler, 2009]. The blood pressure reported in the HUNT study is measured by trained personnel reducing the risk of measurement errors. However, it is not unlikely that errors such as talking, active listening, distant bladder, or smoking within 30 minutes of measurement can have occurred, and according to Handler [2009] this can lead to an error in systolic blood pressure of approximately $10 - 15mmHg$. This error is approximately $7 - 11\%$ of the average systolic blood pressure measured in the HUNT2 and HUNT3 cohorts. Exploring the effect measurement errors have on the models could be of interest in future reaserch.

When developing models and testing the inference, we at times used a smaller dataset and observed that the parameter estimates varied with the size of the dataset. We also observed similar tendencies when creating simulated data for reproducibility purposes that are publicly available. Exploring this discrepancy was considered outside the scope of this thesis. However, this might affect the bias of the SPM when data is MNAR. Future research should contain a thorough analysis of the impact of the data size and how the parameter estimates vary with different data sizes. This research could give a more precise guideline as to when these models are appropriate to use, e.g., how large datasets are needed.

Whether the data is MAR or MNAR relies on untestable assumptions [Enders, 2011]. However, for future research, it would be interesting to study the models and results of this work in the new graphical framework introduced by Mohan and Pearl [2021]. This new approach of missing data modeling could greatly benefit understanding the proposed model's properties.

As a final note, we acknowledge that the models presented in this work are far from trivial to use and are computationally demanding. Especially the use of the software package R-INLA can be challenging for non-statisticians. However, we find the models presented to provide helpful insight into the underlying structure of the HUNT data and how one can model accounting for data MNAR.

To summarize this work, we find that the SPM is better suited to account for data MNAR than the naive model, even though both models are biased. The bias is still unaccounted for but is not caused by misspecification of the informative prior of the association parameter c in the SPM. We have shown that the SPM can be fitted efficiently through INLA and can be used for predictive purposes. In addition, we have proposed a novel scheme for evaluating if data is MNAR when the data follows the SPM. This scheme is tested in simulation studies and indicates that blood pressure missing due to dropout in the HUNT Study is MNAR.

Bibliography

Paul S Albert and Dean A Follmann. Modeling repeated count data subject to informative dropout. *Biometrics*, 56(3):667–677, 2000.

Gunnar H Anderson Jr, Nancy Blakeman, and DH Streeten. The effect of age on prevalence of secondary forms of hypertension in 4429 consecutively referred patients. *Journal of hypertension*, 12(5):609–615, 1994.

Bjørn Olav Åsvold, Arnulf Langhammer, Tommy Aune Rehn, Grete Kjelvik, Trond Viggo Grøntvedt, Elin Pettersen Sørgjerd, Jørn Søberg Fenstad, Oddgeir Holmen, Maria C Stuifbergen, Sigrid Anna Aalberg Vikjord, et al. Cohort profile update: The hunt study, norway. *medRxiv*, 2021.

Haakon Bakka, Håvard Rue, Geir-Arne Fuglstad, Andrea Riebler, David Bolin, Janine Illian, Elias Krainski, Daniel Simpson, and Finn Lindgren. Spatial modeling with r-inla: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6):e1443, 2018.

Narayanaswamy Balakrishnan. *Methods and applications of statistics in the life and health sciences*. John Wiley & Sons, 2009.

Marta Blangiardo and Michela Cameletti. *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons, 2015.

Marta Blangiardo, Michela Cameletti, Gianluca Baio, and Håvard Rue. Spatial and spatio-temporal models with r-inla. *Spatial and spatio-temporal epidemiology*, 4:33–49, 2013.

Glenn W Brier et al. Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3, 1950.

Stephen Brooks. Markov chain monte carlo method and its application. *Journal of the royal statistical society: series D (the Statistician)*, 47(1):69–100, 1998.

Clarice D Brown, Millicent Higgins, Karen A Donato, Frederick C Rohde, Robert Garrison, Eva Obarzanek, Nancy D Ernst, and Michael Horan. Body mass index and the prevalence of hypertension and dyslipidemia. *Obesity research*, 8(9):605–619, 2000.

Thomas A Brown. Admissible scoring systems for continuous distributions. *Santa Monica, California: The Rand Corporation*, 1974.

Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1):1–32, 2017.

An Creemers, Niel Hens, Marc Aerts, Geert Molenberghs, Geert Verbeke, and Michael G Kenward. A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal*, 52(1):111–125, 2010.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

Peter Diggle and Michael G Kenward. Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43(1):49–73, 1994.

Craig K Enders. Missing not at random models for latent growth curve analyses. *Psychological methods*, 16(1):1, 2011.

Lars Fredrik Espeland. A shared parameter model accounting for non-ignorable missing data due to dropout: Modelling of blood pressure based on the hunt study. Master's thesis, Norwegian University of Science and Technology, 7 2020.

Ludwig Fahrmeir, Thomas Kneib, Stefan Lang, and Brian Marx. *Regression*. Springer, 2007.

Dean Follmann and Margaret Wu. An approximate generalized linear model with random effects for informative missing data. *Biometrics*, pages 151–168, 1995.

Ahmed M Gad and Nesma MM Darwish. A shared parameter model for longitudinal data with missing values. *American journal of applied Mathematics and Statistics*, 1(2):30–35, 2013.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

Virgilio Gómez-Rubio. *Bayesian inference with INLA*. Chapman & Hall/CRC Press. Boca Raton, FL., 2020.

Joel Handler. The importance of accurate blood pressure measurement. *The Permanente Journal*, 13(3):51, 2009.

W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika, Volume 57, Issue 1, April 1970, Pages 97–109*, 1970.

James J Heckman. Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 47(1):153–161, 1979.

Robin Henderson, Peter Diggle, and Angela Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.

Hans Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570, 2000.

Aurora Christine Hofman. A-spm-accounting-for-data-mnar, 2021a. URL `https://github.com/AuroraSmil/A-SPM-accounting-for-data-MNAR`.

Aurora Christine Hofman. What about the dropouts? accounting for non-ignorable missing data due to dropouts with a shared parameter model, June 2021b.

Chanelle J Howe, Stephen R Cole, Bryan Lau, Sonia Napravnik, and Joseph J Eron Jr. Selection bias due to loss to follow up in cohort studies. *Epidemiology (Cambridge, Mass.)*, 27(1):91, 2016.

John PA Ioannidis. Why most published research findings are false. *PLoS medicine*, 2(8):e124, 2005.

Shu-Zhong Jiang, Wen Lu, Xue-Feng Zong, Hong-Yun Ruan, and Yi Liu. Obesity and hypertension. *Experimental and therapeutic medicine*, 12(4):2395–2399, 2016.

Alexander Jordan. *Facets of forecast evaluation*. PhD thesis, Karlsruher Institut für Technologie (KIT), 2016.

Alexander Jordan, Fabian Krüger, and Sebastian Lerch. Evaluating probabilistic forecasts with scoringrules. *arXiv preprint arXiv:1709.04743*, 2017.

Niko A Kaciroti and Roderick JA Little. Bayesian sensitivity analyses for longitudinal data with dropouts that are potentially missing not at random: A high dimensional pattern-mixture mode. *Statistics in Medicine*, 40(21):4609–4628, 2021.

S Krokstad, A Langhammer, K Hveem, TL Holmen, K Midthjell, TR Stene, G Bratberg, J Heggland, and J Holmen. Cohort profile: the hunt study, norway. *International journal of epidemiology*, 42(4):968–977, 2013.

Fabian Krüger, Sebastian Lerch, Thordis Thorarinsdottir, and Tilmann Gneiting. Predictive inference based on markov chain monte carlo output. *International Statistical Review*, 89(2): 274–301, 2020.

Lewington et al. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *The Lancet*, 360 (9349):1903–1913, 2002.

Finn Lindgren and Håvard Rue. Bayesian spatial modelling with r-inla. *Journal of statistical software*, 63(1):1–25, 2015.

Antonio R Linero and Michael J Daniels. Bayesian approaches for missing not at random outcome data: The role of identifying restrictions. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 33(2):198, 2018.

Roderick JA Little. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125–134, 1993.

Roderick JA Little. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the american statistical association*, 90(431):1112–1121, 1995.

Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.

Sara Martino and Andrea Riebler. Integrated nested laplace approximations (inla). *Wiley StatsRef: Statistics Reference Online*, pages 1–19, 2019.

Thiago G Martins, Daniel Simpson, Finn Lindgren, and Håvard Rue. Bayesian computing with inla: new features. *Computational Statistics & Data Analysis*, 67:68–83, 2013.

Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

Karthika Mohan and Judea Pearl. Graphical models for processing missing data. *Journal of the American Statistical Association*, 116(534):1023–1037, 2021.

Geert Molenberghs, Caroline Beunckens, Cristina Sotto, and Michael G Kenward. Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2):371–388, 2008.

Christopher JL Murray, Aleksandr Y Aravkin, Peng Zheng, Cristiana Abbafati, Kaja M Abbas, Mohsen Abbasi-Kangevari, Foad Abd-Allah, Ahmed Abdelalim, Mohammad Abdollahi, Ibrahim

Abdollahpour, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet*, 396 (10258):1223–1249, 2020.

Janet van Niekerk, Haakon Bakka, and Håvard Rue. Competing risks joint models using r-inla. *Statistical Modelling*, 21(1-2):56–71, 2021.

Erik P Pulkstenis, Thomas R Ten Have, and J Richard Landis. Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. *Journal of the American Statistical Association*, 93(442):438–450, 1998.

R-INLA. R-inla project, 2021. URL `https://www.r-inla.org/home`.

Eleni Rapsomaniki, Adam Timmis, Julie George, Mar Pujades-Rodriguez, Anoop D Shah, Spiros Denaxas, Ian R White, Mark J Caulfield, John E Deanfield, Liam Smeeth, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1· 25 million people. *The Lancet*, 383(9932):1899–1911, 2014.

Christopher H Rhoads. Problems with tests of the missingness mechanism in quantitative policy studies. *Statistics, Politics, and Policy*, 3(1), 2012.

Donald B Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. CRC press, 2005.

Håvard Rue, Sara Martino, and Nicolas Chopin. 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, 2009.

Håvard Rue, Andrea Riebler, Sigrunn H Sørbye, Janine B Illian, Daniel P Simpson, and Finn K Lindgren. Bayesian computing with inla: a review. *Annual Review of Statistics and Its Application*, 4:395–421, 2017.

David J Spiegelhalter, Andrew Thomas, Nicky Best, and David Lunn. Winbugs version 1.4 user manual. *MRC Biostatistics Unit, Cambridge. URL http://www. mrc-bsu. cam. ac. uk/bugs*, 2003.

Ingelin Steinsland, Camilla Thorrud Larsen, Alexandre Roulin, and Henrik Jensen. Quantitative genetic modeling and inference in the presence of nonignorable missing data. *Evolution*, 68 (6):1735–1747, 2014.

R Thomas, Ten Have, Allen R Kunselman, Erik P Pulkstenis, and J Richard Landis. Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics*, pages 367–383, 1998.

Martin D Tobin, Nuala A Sheehan, Katrina J Scurrah, and Paul R Burton. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Statistics in medicine*, 24(19):2911–2935, 2005.

Masahiko Tozawa, Kunitoshi Iseki, Chiho Iseki, Kozen Kinjo, Yoshiharu Ikemiya, and Shuichi Takishita. Blood pressure predicts risk of developing end-stage renal disease in men and women. *Hypertension*, 41(6):1341–1345, 2003.

Edward F Vonesh, Tom Greene, and Mark D Schluchter. Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in medicine*, 25(1):143–163, 2006.

Paul K Whelton. Epidemiology of hypertension. *Lancet (London, England)*, 344(8915):101–106, 1994.

World Health Organizatoin. Global action plan for the prevention and control of ncds 2013-2020, 2013. URL `https://www.who.int/publications/i/item/9789241506236`.

Margaret C Wu and Raymond J Carroll. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, pages 175–188, 1988.

Bin Zhou, James Bentham, Mariachiara Di Cesare, Honor Bixby, Goodarz Danaei, Melanie J Cowan, Christopher J Paciorek, Gitanjali Singh, Kaveh Hajifathalian, James E Bennett, et al. Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19· 1 million participants. *The Lancet*, 389(10064):37–55, 2017.

A Supplementary Material for The Simulation Studies on Bias and Coverage

We have summarized the results from the simulation studies performed in Section 5.1 exploring the bias and coverage of the SPM (3.3) and the naive model (3.4). Table A.1, Table A.2, Table A.3, and Table A.4 display the mean posterior mean, bias, and coverage of the parameter estimates for both the SPM and the naive model. Further we display the difference in bias for the SPM and naive model $Bias_{SPM} - Bias_{naive}$.

Table A.1.: A summary of the results from simulation study 1 Table 5.1 where we explore the bias and coverage of the SPM (3.3) and the naive model (3.4) when the true model parameters are known. We display the mean posterior mean, bias of posterior means, and coverage for all parameters. In addition, we display the difference in bias for the SPM and the naive model ($Bias_{SPM} - Bias_{naive}$). The true parameters are the posterior means from the SPM (3.3) as given in Table 4.1 and the data is MNAR.

		SPM			**Naive model**			
	True value	Mean	Bias	Coverage	Mean	Bias	Coverage	Difference in bias
α_0	0.27	0.24	-3e-02	0.07	0.13	-0.144	0.00	0.112
α_{BP}	0.60	0.60	-4e-03	0.91	0.58	-0.021	0.01	0.017
α_{age}	0.24	0.24	-8e-04	0.98	0.25	0.002	0.97	0.001
α_{bmi}	0.07	0.07	-1e-03	0.95	0.07	-0.006	0.76	0.004
α_{sex}	0.04	0.04	-3e-03	0.97	0.02	-0.018	0.42	0.015
β_0	0.57	0.56	-1e-02	0.98	0.55	-0.016	0.99	0.004
β_{BMI}	0.09	0.13	5e-02	0.01	0.13	0.042	0.01	-0.005
β_{sex}	0.29	0.29	-5e-03	0.93	0.28	-0.016	0.86	0.011
β_{BP}	0.14	0.14	-3e-03	0.90	0.13	-0.007	0.91	0.004
σ_{age}	1.42	1.22	-2e-01	0.67	1.34	-0.082	1.00	-0.119
σ_ϵ	0.79	0.78	-8e-03	0.32	0.77	-0.020	0.00	0.012
c	0.70	0.55	-2e-01	0.00	NA	NA	NA	NA

Table A.2.: A summary of the results from simulation study 2 Table 5.1 where we explore the bias and coverage of the SPM (3.3) and the naive model (3.4) when the true model parameters are known. We display the mean posterior mean, bias of posterior means, and coverage for all parameters. Further, we display the difference in bias for the SPM and the naive model ($Bias_{SPM} - Bias_{naive}$).The true parameters are the posterior means from the SPM (3.3) as given in Table 4.1 and the data is MAR meaning $c = 0$.

		SPM			**Naive model**			
	True value	Mean	Bias	Coverage	Mean	Bias	Coverage	Difference in bias
α_0	0.27	0.272	-3e-03	0.8	0.27	-2e-04	1.0	-3e-03
α_{BP}	0.60	0.599	-8e-04	1.0	0.60	-4e-04	1.0	-4e-04
α_{age}	0.24	0.243	-6e-04	1.0	0.24	-7e-04	1.0	1e-04
α_{bmi}	0.07	0.074	6e-04	0.9	0.07	7e-04	0.9	1e-04
α_{sex}	0.04	0.041	7e-05	0.9	0.04	4e-04	1.0	4e-04
β_0	0.57	0.553	-1e-02	1.0	0.57	-2e-03	1.0	-1e-02
β_{BMI}	0.09	0.139	5e-02	0.0	0.14	5e-02	0.0	2e-04
β_{sex}	0.29	0.292	-4e-05	0.9	0.29	9e-05	0.9	4e-05
β_{BP}	0.14	0.139	3e-04	0.9	0.14	2e-04	0.9	-1e-04
σ_{age}	1.42	1.069	-4e-01	0.4	1.41	-6e-03	1.0	-3e-01
σ_ϵ	0.79	0.790	4e-04	1.0	0.79	3e-04	1.0	-1e-04
c	0.00	-0.008	-8e-03	0.9	NA	NA	NA	NA

Table A.3.: Summary of the results from simulation study 3 Table 5.1 where we explore the bias and coverage of the SPM (3.3) and the naive model (3.4) when the true model parameters are known. We display the mean posterior mean, bias of posterior means, and coverage for all parameters. In addition, we display the difference in bias for the SPM and the naive model ($Bias_{SPM} - Bias_{naive}$). In addition, we display the difference in bias for the SPM and the naive model $Bias_{SPM} - Bias_{naive}$. The true parameters are the posterior means from the naive model (3.4) as given in Table 4.1 and the data is MNAR.

		SPM			**Naive model**			
	True value	Mean	Bias	Coverage	Mean	Bias	Coverage	Difference in bias
α_0	0.13	0.10	-3e-02	0.05	-0.005	-0.139	0.00	1e-01
α_{BP}	0.58	0.57	-4e-03	0.92	0.558	-0.020	0.04	2e-02
α_{age}	0.25	0.25	-2e-04	0.91	0.248	0.001	0.90	9e-04
α_{bmi}	0.07	0.07	-5e-04	0.98	0.066	-0.005	0.85	4e-03
α_{sex}	0.02	0.02	-1e-03	0.95	0.011	-0.014	0.60	1e-02
β_0	0.55	0.53	-2e-02	0.96	0.527	-0.025	0.96	5e-03
β_{BMI}	0.08	0.13	5e-02	0.00	0.127	0.045	0.01	-5e-03
β_{sex}	0.27	0.27	-2e-03	0.94	0.262	-0.013	0.87	1e-02
β_{BP}	0.13	0.13	-3e-03	0.91	0.126	-0.007	0.90	4e-03
σ_{age}	1.40	1.29	-1e-01	0.62	1.340	-0.061	1.00	-5e-02
σ_ϵ	0.77	0.76	-8e-03	0.33	0.750	-0.020	0.00	1e-02
c	0.70	0.55	-2e-01	0.02	NA	NA	NA	NA

Table A.4.: Summary of the results from simulation study 4 Table 5.1 where we explore the bias and coverage of the SPM (3.3) and the naive model (3.4) when the true model parameters are known. We display the mean posterior mean, bias of posterior means, and coverage for all parameters. Further, we display the difference in bias for the SPM and the naive model ($Bias_{SPM} - Bias_{naive}$). The true parameters are the posterior means from the naive model (3.4) as given in Table 4.1 and the data is MAR meaning $c = 0$.

		SPM			**Naive model**			
	True value	Mean	Bias	Coverage	Mean	Bias	Coverage	Difference in bias
α_0	0.13	0.14	1e-03	0.8	0.13	-7e-04	0.9	-4e-04
α_{BP}	0.58	0.58	5e-04	0.9	0.58	2e-04	1.0	-3e-04
α_{age}	0.25	0.25	9e-04	0.9	0.25	1e-03	0.9	4e-05
α_{bmi}	0.07	0.07	3e-04	1.0	0.07	2e-04	1.0	-8e-05
α_{sex}	0.02	0.03	1e-03	0.9	0.03	7e-04	0.9	-3e-04
β_0	0.55	0.53	-2e-02	0.9	0.55	-7e-03	1.0	-1e-02
β_{BMI}	0.08	0.13	5e-02	0.0	0.13	5e-02	0.0	2e-04
β_{sex}	0.27	0.28	1e-03	0.9	0.28	1e-03	1.0	1e-04
β_{BP}	0.13	0.13	6e-04	0.9	0.13	4e-04	0.9	-1e-04
σ_{age}	1.40	0.94	-5e-01	0.3	1.40	-5e-04	1.0	-5e-01
σ_ϵ	0.77	0.77	-3e-05	0.9	0.77	-7e-05	1.0	3e-05
c	0.00	0.01	1e-02	0.9	NA	NA	NA	NA

B Shared Parameter and Naive Model Results on Simulated Data

For comparison concerning reproducibility purposes, we fitted our main SPM (3.3) and naive model (3.4) to simulated data created as described in Chapter 7. In the GitHub repository by Hofman [2021a] two simulated datasets are available, one with 1000 and one with 64385 simulated participants. We see from Figure B.1 and Figure B.2 that we obtain similar, although not identical, results for the model fits on the simulated dataset of the same size as the HUNT2 cohort as on the HUNT2 cohort. When we use the smaller simulated dataset with only 1000 participants, we do not get the same results as for the real data. This can be seen in Figure B.3 and Figure B.4. We, therefore, note that while the smaller dataset can be used to verify the code, it can not be used to compare results. Further, we note that this indicates that a certain amount of data is needed for using the models presented in this work.

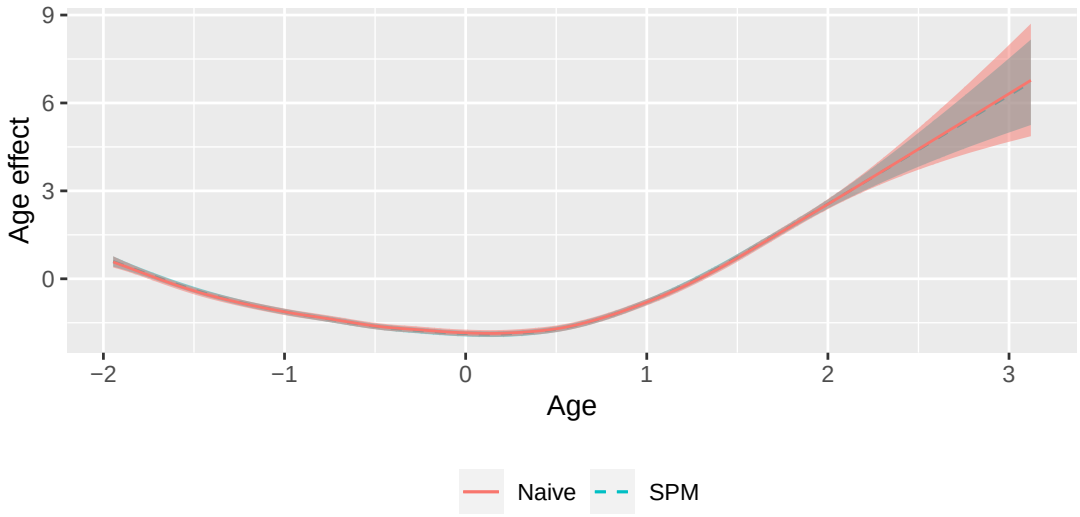


Figure B.1.: Age effect for the SPM and naive model fitted to simulated data mimicking the size (64385 participants) and structure of the HUNT2 cohort with 95% credible bands.

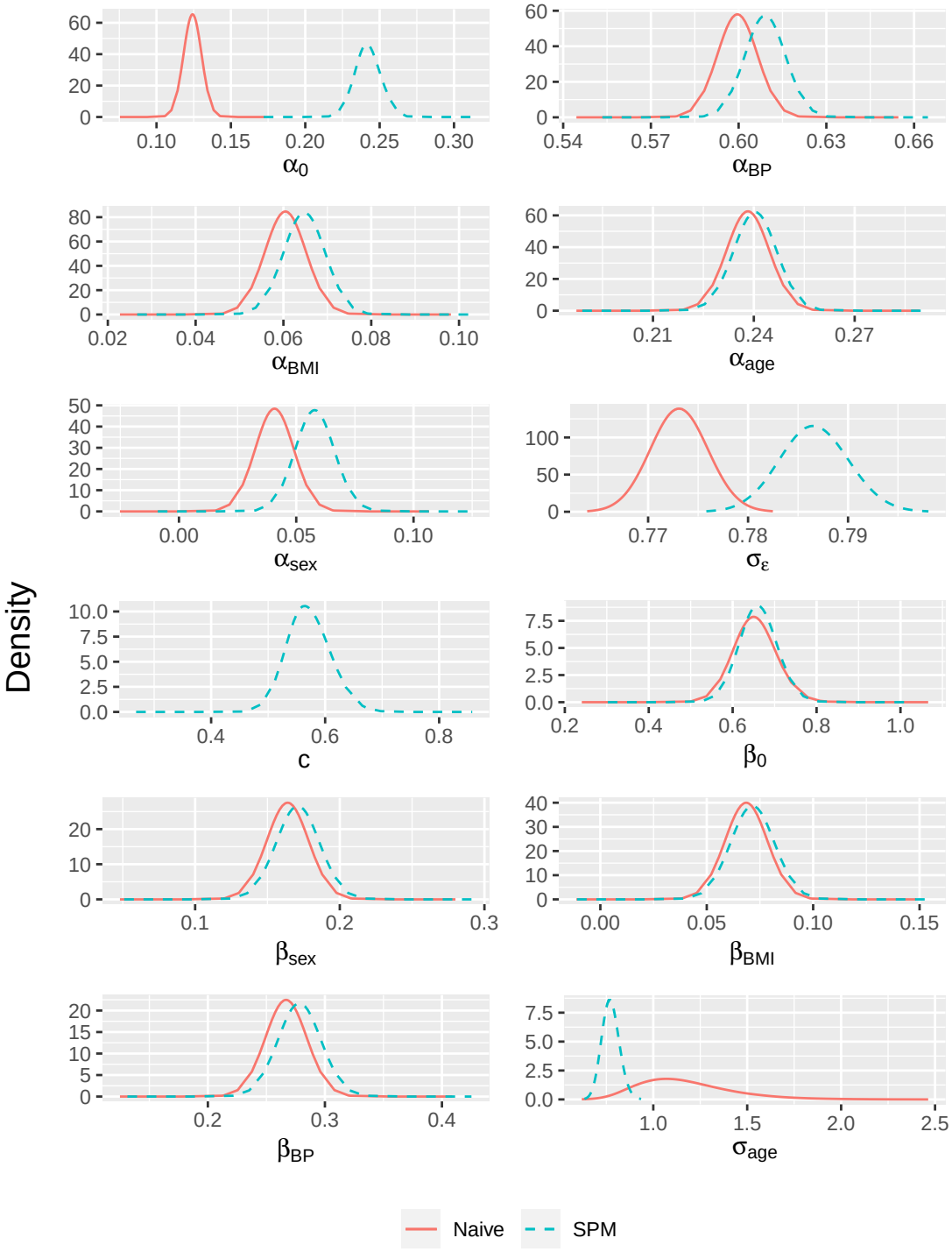


Figure B.2.: Posterior distribution of the latent field and hyperparameters for the SPM (3.3) and naive model (3.4) fitted to simulated data mimicking the size (64385 participants) and structure of the HUNT2 cohort.

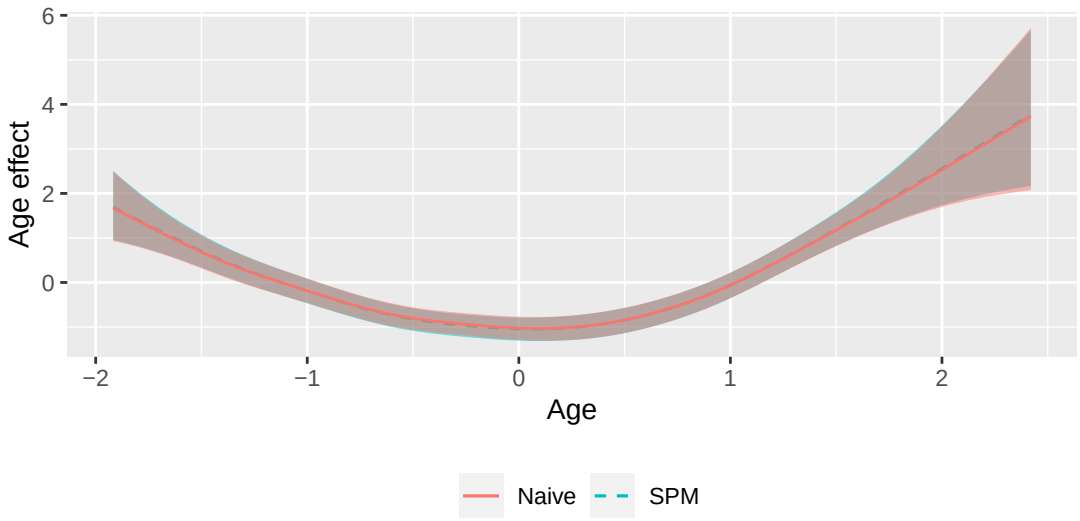


Figure B.3.: Age effect for the SPM and naive model fitted to a small simulated dataset (1000 participants) mimicking the structure of the HUNT2 cohort with 95% credible bands.

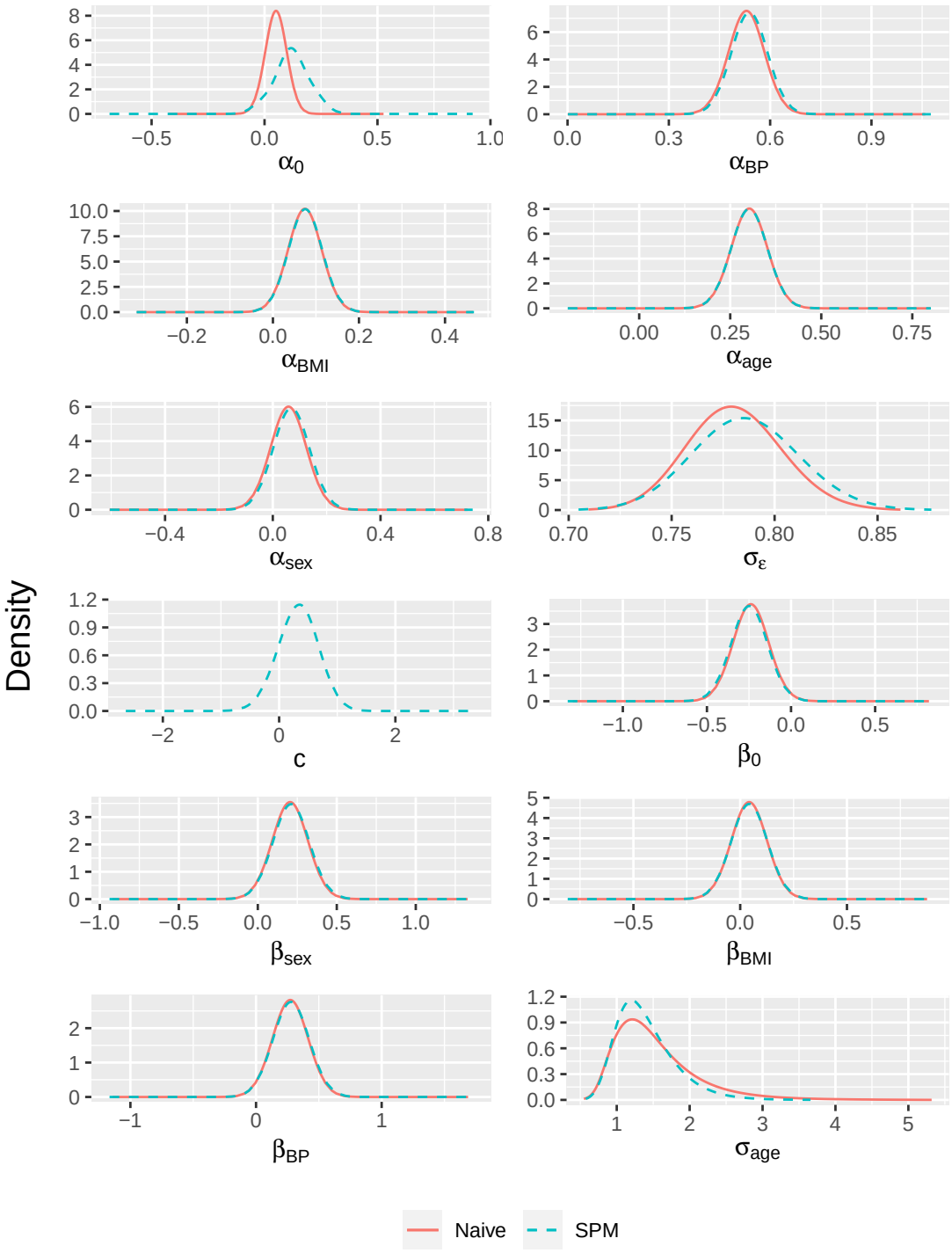


Figure B.4.: Posterior distribution of the latent field and hyperparameters for the SPM (3.3) and naive model (3.4) fitted to a small simulated dataset (1000 participants) mimicking the structure of the HUNT2 cohort.

Papers

C

During the time of writing this thesis, Ingelin Steinsland, Lars Espeland, Emma Ingeström, and I wrote a small paper based on the work by Espeland [2020] and a full journal paper based on the work by Espeland [2020], my project thesis Hofman [2021b], and this master thesis.

The first, "A Shared parameter model accounting for dropout not at random in a predictive model for systolic blood pressure using The HUNT Study", was submitted to the 2021 Neural Information Processing Systems conference (14. December 2021) and accepted, after being peer-reviewed, for a poster presentation.

The second "A Shared Parameter Model for Systolic Blood Pressure Accounting for Data Missing Not at Random in the HUNT Study" is finalized at the cusp of completing this thesis and will shortly be sent to the HUNT team for approval before it is submitted to an international recognized journal in statistics. The manuscript given here is not considered complete for submission when delivering this work. However, it is included here for completeness. It shows an essential milestone in working with this thesis, and a considerable amount of effort went into it.

Both papers are self-contained.

A Shared parameter model accounting for dropout not at random in a predictive model for systolic blood pressure using The HUNT Study

Aurora C. Hofman
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim
aurorach@stud.ntnu.no

Lars Espeland
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim
larsespeland1@gmail.com

Ingelin Steinsland
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim
ingelin.steinsland@ntnu.no

Emma M. L. Ingeström
Department of Circulation and Medical Imaging
Norwegian University of Science and Technology
Trondheim
emma.ingestrom@ntnu.no

Abstract

This work proposes and evaluates a shared parameter model (SPM) to account for data being missing not at random (MNAR) for a model based on a longitudinal population study. The aim is to model systolic blood pressure ten years ahead based on current observations. It is inspired by and evaluated on data from the Trøndelag Health Study (HUNT). The proposed SPM consists of a linear model for the systolic blood pressure and a logistic model for the dropout process connected through a shared random effect. To evaluate the SPM we compare the parameter estimates of the SPM with a naive linear Bayesian model using the same explanatory variables while ignoring the dropout process. This corresponds to assuming data to be missing at random (MAR). In addition, a simulation study is performed in which the naive model and the SPM are tested on data with known parameters when missingness is assumed to be MNAR. The SPM indicates that participants with higher systolic blood pressure than expected from the explanatory variables at the time of the follow-up study have a higher probability of dropping out, suggesting that the data are MNAR. Further, the SPM and the naive model result in different parameter estimates for the explanatory variables.

NeurIPS 2021 Workshop Your Model Is Wrong: Robustness and Misspecification in Probabilistic Modeling.

1 Introduction

An elevated systolic blood pressure (BP) increases the risk of developing diseases related to the heart, blood vessels, brain, and kidney [Lewington et al., 2002, Rapsomaniki et al., 2014, Tozawa et al., 2003]. Elevated BP affects over 1.1 billion people worldwide and accounts for 10.8 million global deaths per year [Zhou et al., 2017, Murray et al., 2020]. Early detection, treatment, and control of elevated BP are of high priority in public health strategies [World Health Organization, 2013]. Hence, the ability to predict future BP trends and use this to start preventive measures early is of great benefit [Whelton, 1994]. This work is motivated by making models for understanding BP ten years ahead based on current BP, age, and sex using data from Helseundersøkelsen i Trøndelag (HUNT). The HUNT Study is a series of four comprehensive population-based health surveys in central Norway and a valuable source for longitudinal research. However, longitudinal cohort studies following participants over a long period may be subject to selection bias as participants being lost to follow-up, i.e., dropout between two subsequent surveys, cause data to be missing. The missing data may contain important information about the true distribution of the data and must therefore be treated with great care [Little and Rubin, 2019].

This work proposes a shared parameter model (SPM) assuming data missing not at random (MNAR) for BP based on a population-based health study [Wu and Carroll, 1988, Follmann and Wu, 1995, Vonesh et al., 2006, Steinsland et al., 2014]. The model fit belongs to the framework of a Bayesian latent Gaussian model, and computational efficient inference is available through integrated nested Laplace approximations (INLA)[Rue et al., 2009, 2017, Gómez-Rubio, 2020]. Further, the SPM is compared to a naive model under the assumption of data being missing at random (MAR). We use the definitions as stated in Little and Rubin [2019]. The differences between these categories are how the missingness depends on the complete data. Data MAR means the distribution of missingness is independent of the missing response values but can be dependent on the observed data. Data MNAR means the probability of being missing depends on the missing values themself. This process is non-ignorable and needs to be modeled together with the original model [Little and Rubin, 2019]. The work presented here is based on the thesis by [Espeland, 2020].

2 Data

Summary of data from HUNT1-HUNT2

	Summary HUNT1	Observed in HUNT2	Missing in HUNT2
Variable	n = 74 247	n = 46 456, 62.6 %	n = 27 791, 37.4 %
BP_2	-	141.66 ± 22.62	-
BP_1	138.48 ± 23.52	133.66 ± 19.94	146.54 ± 26.65
age_1	49.30 ± 17.45	45.17 ± 14.24	56.20 ± 19.98
sex			
female	37 788	24 793 (65.6 %)	12 995 (34.4 %)
male	36 459	21 663 (59.5 %)	14 796 (40.5 %)

Table 1: Summary of the HUNT1-HUNT2 cohort. The second column displays the mean and standard deviation for all continuous variables of all participants in HUNT1 in addition to the number of females/males present in the study. The third and fourth column displays the number of participants from HUNT1 who participated or dropped out in HUNT2 and the mean and variance of all continuous variables of the participants grouped by missingness in HUNT2. They also display the percentage of the original number of females/males who participated/dropped out. All the data are displayed on the original data scale.

In this work, we follow a cohort of adults participating in HUNT1 (1984-86) until participation in HUNT2 (2006-08). Of 74 247 participants with complete records on BP, age, sex, and BMI at HUNT1, 27 791 (37.4 %) participants were missing in HUNT2 meaning the response value of interest is missing in a large proportion of the sample [Krokstad et al., 2013]. The available data do not contain any information on the reason for dropout, e.g., declined to participate, moved out of the county, or being dead. A summary of the data from the HUNT Study is provided in Table 1. The indices in BP1, age1, and BP2 refer to age and BP from HUNT1 and HUNT2 respectively. Prior to modeling the explanatory variables are standardized based on the empirical mean and variance from HUNT1.

3 Models and methods

3.1 Naive and shared parameter model

The shared parameter model consists of two sub-models, one for BP (BP_2) and one for the dropout process (m_2). The index i signifies participant number i. The first model is given by,

$$BP_{2i} = \alpha_0 + \alpha_{BP} BP_{1i} + \alpha_{age} age_{1i} + \alpha_{sex} sex_i + \epsilon_i. \tag{3.1}$$

This sub-model is the naive model when the dropout process is ignored. When the dropout process is accounted for ϵ_i is shared with the sub-model for the dropout process given as follows,

$$\text{logit}(p_{2i}) \sim \beta_0 + \beta_{BP} BP_{1i} + f(age_{1i}) + \beta_{sex} sex_i + c\epsilon_i \tag{3.2}$$

$$m_{2i} \sim \text{Bernoulli}(p_{2i}). \tag{3.3}$$

Here p_{2i} is the probability of dropping out for participant i, $f(.)$ is a random walk of order 2 [Gómez-Rubio, 2020], and the parameter c is an association parameter that links the two models together. A random walk is a common choice to include smooth terms in the linear predictor [Gómez-Rubio, 2020]. As shown by Espeland [2020] only age has an additive effect even when all continuous parameters are modeled non-linearly in the naive model. In addition adding additive effects severely decreases the computational speed. Hence, as suggested by Espeland [2020] we only model age as a non-linear effect. The variances of the second-order difference for the random walk and the variance for the individual random effect are assigned a non-informative Gamma distribution,

$$\sigma_{age}, \sigma_\epsilon \sim \text{Gamma}(1, 5 \cdot 10^5). \tag{3.4}$$

The priors for the coefficients are,

$$\alpha_0, \alpha_{BP}, \alpha_{age}, \alpha_{sex}, \beta_0, \beta_{BP}, \beta_{sex} \sim N(0, 10^6). \tag{3.5}$$

The association-parameter c is given a more informative normal prior,

$$c \sim N(0, 1) \tag{3.6}$$

as this parameter is assumed to be small.

3.2 Inference

By modeling the BP in a Bayesian framework we obtain the posterior marginals of all the coefficients involved. This can be done using integrated nested Laplace approximations (INLA) because the SPM and the naive model are parts of the class of latent Gaussian models [Rue et al., 2009, 2017, Gómez-Rubio, 2020].

We use the R-INLA software (`https://www.r-inla.org`) which also supports fitting models with multiple likelihoods, which is the case for the model specified in Equation 3.1 and 3.2 where the dropout process is modeled with a binomial likelihood and the model for BP is modeled with a normal likelihood.

3.3 Simulation study

To evaluate the performance of the models a simulation study is performed under MNAR assumption. The parameters are the posterior mean estimates obtained by the SPM for HUNT1-HUNT2 data (Table 2). These are used to simulate BP and the dropout process between HUNT1 and HUNT2.

We simulate $j = 1 : 100$ new data sets and fit both the naive model and the SPM to the data. To compare the models both the mean posterior mean, mean bias and coverage are compared. The mean bias is given by, $\text{Bias}(\hat{\theta}) = \sum_{j=1}^{j=100} (\hat{\theta_j} - \theta)$ The coverage is defined as the proportion of occurrences where the true parameter θ falls within the 95% equal-tailed credible interval.

4 Results

The results of the parameter estimations can be seen in Figure 1. We can see that all coefficients but α_{sex} in the SPM are significant in both models and do not have zero contained in their distribution.

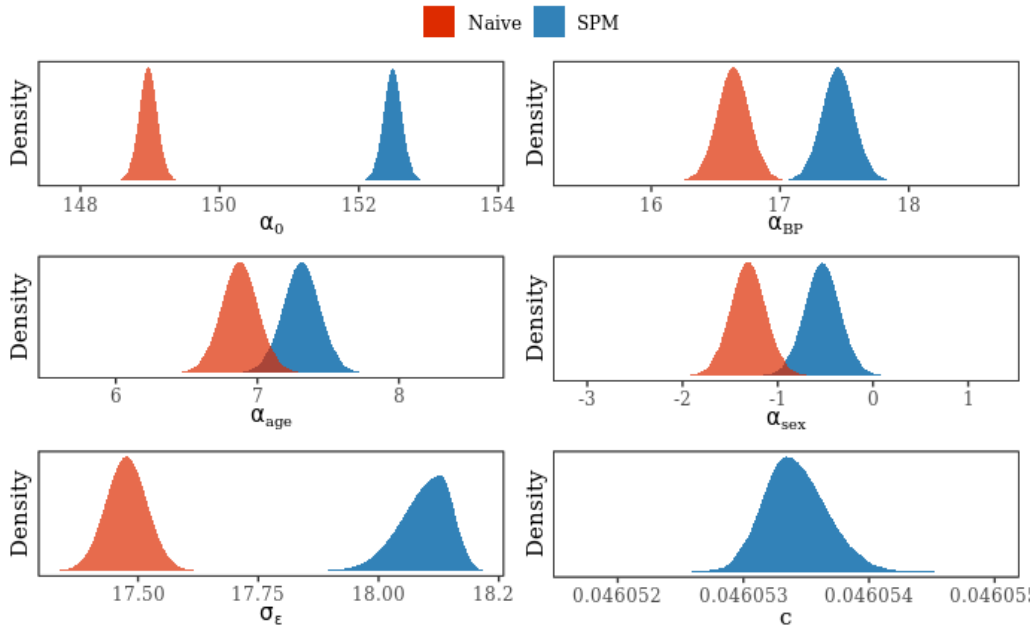


Figure 1: Posterior density for the parameters and hyperparameters of both the SPM model and the naive model for the BP model (3.1). We used the HUNT1-HUNT2 data as described in Table 1 with scaled explanatory variables. The association parameter c (3.2) in the SPM is also displayed.

We also note that the posterior mean for the association parameter, c (3.2), is unequal to zero which can also be seen in the bottom right panel of Figure 1.

When comparing the estimates made by the SPM with the naive model we see that the naive model estimates a lower effect of the explanatory variables than the SPM for all explanatory variables. From Figure 1 we also see that the coefficients with the smallest overlap in posterior density between the SPM and the naive model are α_0, α_{BP_1}, and σ_ϵ. The estimates of α_{age} and α_{sex} have more overlap although still distinct.

		SPM			**Naive**		
	True	Mean	Bias	Coverage	Mean	Bias	Coverage
α_0	152.49	151.50	-0.99	0.14	148.94	-3.55	0
α_{BP_1}	17.45	17.23	-0.22	0.49	16.65	-0.80	0
α_{age_2}	7.31	7.20	-0.11	0.82	6.94	-0.37	0.13
α_{sex}	-0.53	-0.75	-0.22	0.71	-1.35	-0.82	0
β_0	0.13	0.11	-0.02	0.96	-	-	-
β_{BP_1}	0.26	0.25	-0.01	0.79	-	-	-
β_{sex}	0.44	0.42	-0.02	0.75	-	-	-
σ_ϵ	18.10	17.70	-0.40	0.12	-17.45	-0.65	0.03
c	0.046	0.033	-0.013	0.07	-	-	-

Table 2: Summary of the parameter estimates obtained from the simulation study when the data are MNAR. The mean and biases are the mean of the posterior means and the bias of the posterior means as defined in Section 3.3. The coverage is defined as the proportion of times the true parameter value falls within the 95% equally tailed credible interval.

The results of the simulation study are summarized in Table 2. The SPM has relatively good parameter estimates and overall good coverage. We see that the SPM estimates all parameters closer to the true value than the naive model. The coverage varies much between the parameters, but the SPM has consistently better coverage than the naive model. The low coverage for c could be due to prior sensitivity. We also see that both the SPM and the naive model underestimate the true parameter

values for all parameters. However, the bias is clearly reduced by the SPM compared to the naive model in our simulation.

5 Closing Remarks

Models accounting for data MNAR are based on untestable model assumptions. As stated by Molenberghs et al. [2008] "each MNAR model fit to a set of observed data can be reproduced exactly by a MAR counterpart". In this work, we suggest reasonable models for both blood pressure and the missing process, and the results indicate data MNAR that needs to be considered. The posterior distribution of the association parameter in the SPM has all its mass above zero and the SPM and the naive model result in different posterior distributions. The simulation study confirms that if the underlying assumption of the data being MNAR holds and a higher BP ten years ahead than explained by the explanatory variables increases the probability of dropping out, the SPM gives less bias and better coverage. Further, we demonstrate that the suggested SPM can be fitted using INLA, and is computationally feasible for a data set of this size. Future work should include more simulation studies including data MAR, a thorough sensitivity analysis, and different scaling of the data to possibly solve numerical issues.

References

Lars Fredrik Espeland. A shared parameter model accounting for non-ignorable missing data due to dropout: Modelling of blood pressure based on the hunt study. Master's thesis, Norwegian University of Science and Technology, 7 2020.

Dean Follmann and Margaret Wu. An approximate generalized linear model with random effects for informative missing data. *Biometrics*, pages 151–168, 1995.

Virgilio Gómez-Rubio. *Bayesian inference with INLA*. Chapman & Hall/CRC Press. Boca Raton, FL., 2020.

S Krokstad, A Langhammer, K Hveem, TL Holmen, K Midthjell, TR Stene, G Bratberg, J Heggland, and J Holmen. Cohort profile: the hunt study, norway. *International journal of epidemiology*, 42 (4):968–977, 2013.

Lewington et al. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *The Lancet*, 360(9349): 1903–1913, 2002.

Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.

Geert Molenberghs, Caroline Beunckens, Cristina Sotto, and Michael G Kenward. Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2):371–388, 2008.

Christopher JL Murray, Aleksandr Y Aravkin, Peng Zheng, Cristiana Abbafati, Kaja M Abbas, Mohsen Abbasi-Kangevari, Foad Abd-Allah, Ahmed Abdelalim, Mohammad Abdollahi, Ibrahim Abdollahpour, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet*, 396(10258): 1223–1249, 2020.

Eleni Rapsomaniki, Adam Timmis, Julie George, Mar Pujades-Rodriguez, Anoop D Shah, Spiros Denaxas, Ian R White, Mark J Caulfield, John E Deanfield, Liam Smeeth, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1· 25 million people. *The Lancet*, 383(9932):1899–1911, 2014.

Håvard Rue, Sara Martino, and Nicolas Chopin. 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, 2009.

Håvard Rue, Andrea Riebler, Sigrunn H Sørbye, Janine B Illian, Daniel P Simpson, and Finn K Lindgren. Bayesian computing with inla: a review. *Annual Review of Statistics and Its Application*, 4:395–421, 2017.

Ingelin Steinsland, Camilla Thorrud Larsen, Alexandre Roulin, and Henrik Jensen. Quantitative genetic modeling and inference in the presence of nonignorable missing data. *Evolution*, 68(6): 1735–1747, 2014.

Masahiko Tozawa, Kunitoshi Iseki, Chiho Iseki, Kozen Kinjo, Yoshiharu Ikemiya, and Shuichi Takishita. Blood pressure predicts risk of developing end-stage renal disease in men and women. *Hypertension*, 41(6):1341–1345, 2003.

Edward F Vonesh, Tom Greene, and Mark D Schluchter. Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in medicine*, 25(1):143–163, 2006.

Paul K Whelton. Epidemiology of hypertension. *Lancet (London, England)*, 344(8915):101–106, 1994.

World Health Organization. Global action plan for the prevention and control of ncds 2013-2020, 2013. URL `https://www.who.int/publications/i/item/9789241506236`.

Margaret C Wu and Raymond J Carroll. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, pages 175–188, 1988.

Bin Zhou, James Bentham, Mariachiara Di Cesare, Honor Bixby, Goodarz Danaei, Melanie J Cowan, Christopher J Paciorek, Gitanjali Singh, Kaveh Hajifathalian, James E Bennett, et al. Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19· 1 million participants. *The Lancet*, 389(10064):37–55, 2017.

Acknowledgements

The Trøndelag Health Study (HUNT) is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology NTNU), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health.

A Shared Parameter Model for Systolic Blood Pressure Accounting for Data Missing Not at Random in the HUNT Study.

Aurora Christine Hofman
Department of Mathematical Sciences
NTNU (Norwegian University of Science and Technology)
and
Lars Espeland
Department of Mathematical Sciences
NTNU (Norwegian University of Science and Technology)
and
Ingelin Steinsland
Department of Mathematical Sciences
NTNU (Norwegian University of Science and Technology)
and
Emma M. L. Ingeström
Department of Circulation and Medical Imaging
NTNU (Norwegian University of Science and Technology)

January 28, 2022

Abstract

In this work, blood pressure eleven years ahead is modeled using data from a longitudinal population-based health survey, the Trøndelag Health (HUNT) Study, while accounting for missing data due to dropout. In the HUNT Study, $20-50\%$ drop out between consecutive surveys. We propose and validate a shared parameter model (SPM) in the Bayesian framework with age, sex, body mass index, and initial blood pressure as explanatory variables. Further, we propose a novel evaluation scheme to evaluate if data is missing not at random (MNAR) based on comparing the predictive performance of the fitted SPM with and without conditioning on the missing process. The results demonstrate that the SPM is suitable for inference for a dataset of this size (cohort of 64385 participants) and structure and indicates data MNAR. The SPM gives different parameter estimates than a naive model assuming data missing at random. The SPM and naive models are compared based on predictive performance

for the validation dataset. The naive model performs slightly better than the SPM for the present participants. However, we find that the naive model performs better for the present participants in a simulation study based on the SPM, while the SPM performs better for the dropouts.

Keywords: longitudinal studies, missing data, INLA (integrated nested Laplace approximations), dropout, health survey

1 Introduction

This work aims to establish and validate a predictive model for systolic blood pressure using data from a longitudinal population-based health survey, the Trøndelag Health (HUNT) Study, while accounting for missing data due to dropout.

Elevated blood pressure increases the risk of developing diseases related to the brain, heart, blood vessels, and kidney [Lewington et al., 2002, Tozawa et al., 2003, Rapsomaniki et al., 2014]. It affects more than 1.1 billion people and accounts for over 10.8 million deaths per year, thereby surpassing smoking as the leading preventable cause of death for middle-aged and older adults worldwide [Zhou et al., 2017, Murray et al., 2020]. Early detection, prevention, and treatment of elevated blood pressure are of high priority in public health strategies [World Health Organizatoin, 2013]. Thus, obtaining unbiased, accurate models for predicting future blood pressure is of great interest in medical research [Whelton, 1994].

The HUNT Study follows the cohort study design meaning the same participants are followed over a long period. In each survey there are between 50000 and 80000 participants, and between consecutive health surveys, $20 - 50\%$ of participants are lost to follow-up [Krokstad et al., 2013, Åsvold et al., 2021]. Proper handling of missing data is vital to obtain unbiased inference [Gad and Darwish, 2013, Little and Rubin, 2019, Chap. 1.3, 6], and how to handle missing data depends on the missing process.

Based on available literature [Anderson Jr et al., 1994, Whelton, 1994, Brown et al., 2000, Jiang et al., 2016, Espeland, 2020] and available observations in the HUNT Study, we suggest a predictive model of future blood pressure with age, sex, body mass index (BMI), and initial blood pressure as explanatory variables. All participants have full records for the explanatory variables. Missing data can be categorized and described in terms of three missing processes; missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) [Little and Rubin, 2019, Chap. 1.3, 6]. If the probability of drop out, i.e. of missing future blood pressure values, is independent of all observed and unobserved data, including the missing response variables and all explanatory variables, the data is MCAR. It is reasonable that the probability of dropping depends on age, and hence MCAR is disregarded. The data is MAR if the probability of missingness dependends on the observed data, but is independent of the unobserved data. If all explanatory variables are observed a model for the missing process with the same explanatory variables as for blood pressure, is an example of a model assuming MAR. Missing processes that are MCAR or MAR are ignorable, meaning unbiased inference can be performed without modeling the missing process. Data that is neither MCAR nor MAR is MNAR [Gad and Darwish, 2013, Little and Rubin, 2019, Chap. 6]. If the part of the future blood pressure which can not be explained by age, sex, BMI, and initial blood pressure, affects the probability of dropping out, the data is MNAR. This can be thought of as (unknown) explanatory variables not

included in the models. It is reasonable to assume that there are health related variables that influence both blood pressure and the probability of drop out. Thus, we argue that in a predictive model for future blood presure we should consider that data might be MNAR. If data is MNAR the missing process must be modeled simultaneously with the original model to obtain unbiased inference [Little and Rubin, 2019, Chap. 1.3, 6].

Even though the assumption of data MAR is often not fulfilled, many of the available software packages and methods described in the literature assume data to be MAR [Balakrishnan, 2009, Rhoads, 2012, Little and Rubin, 2019, Mohan and Pearl, 2021]. However, several studies, especially in biostatistics, have accounted for missing data under the assumption of data MNAR. [Wu and Carroll, 1988, Little, 1993, Diggle and Kenward, 1994, Follmann and Wu, 1995, Little, 1995, Albert and Follmann, 2000, Molenberghs et al., 2008, Howe et al., 2016]. Popular choices for models accounting for data MNAR include the pattern mixture model, selection model, and shared parameter model (SPM) [Heckman, 1979, Wu and Carroll, 1988, Little, 1993, Henderson et al., 2000, Linero and Daniels, 2018, Little and Rubin, 2019, Chap. 15.4]. The SPM is based on the idea of a commonly shared variable affecting both the measurement process and the missing process. Given this variable, the two marginal densities are conditionally independent. It has been used to model longitudinal data subject to MNAR in several studies [Wu and Carroll, 1988, Follmann and Wu, 1995, Thomas et al., 1998, Pulkstenis et al., 1998, Vonesh et al., 2006, Creemers et al., 2010]. In this work, we propose a Bayesian SPM for future blood pressure. The model fits the framework of Bayesian latent Gaussian model and is suitable for Bayesian inference using computationally efficient Integrated Nested Laplace Approximations (INLA) [Rue et al., 2009, 2017, Martino and Riebler, 2019, Gómez-Rubio, 2020, Steinsland et al., 2014].

Molenberghs et al. stated that "each MNAR model fit to a set of observed data can be reproduced exactly by a MAR counterpart" [Molenberghs et al., 2008, p. 371]. Hence, the choice between models eventually comes down to choosing the most likely model assumptions [Enders, 2011]. Recent research has proven that taking the approach of causal modeling and formulating the models through missingness graphs can give theoretical understanding and asymptotic performance guarantees [Mohan and Pearl, 2021]. To the best of our knowledge, the literature provides little insight into practical validation of model performance on data MNAR. The current standard seems to be the use of simulation studies to check the reproducibility of the model. i.e., how well the original parameters are reproduced on simulated data, and sensitivity analysis to check the robustness of the models [Enders, 2011, Steinsland et al., 2014, Kaciroti and Little, 2021].

In this work, we validate the models on a validation dataset. First, predictive performance of the SPM and a naive model assuming the data to be MAR are compared based on the proper scoring rules [Gneiting and Raftery, 2007] continuous ranked probability score (CRPS) and Brier score. Second, we propose a new method to evaluate if data is MNAR based on the SPM. The key idea is that if data is MNAR the missing status has information about the quantity of interest. Therefore we compare the predictive performance of the SPM with and without conditioning on missing status.

The main contributions of this paper is the SPM for blood pressure based on data from the HUNT Study, the demonstration of the applicability for a large case study together with the new insight from the proposed validation schemes.

Section 2 provides background about latent Gaussian models and missing data theory. Section 3 introduces the blood pressure case study including the HUNT Study, the proposed

models, and methods for inference and validation. The results from the case study are presented in Section 4. Section 5 consists of several simulation sensitivity studies based on the HUNT Study. Section 6 summarizes and discusses our findings.

2 Background Theory

This section briefly introduces needed background on latent Gaussian models, and commonly used models and methods for missing data.

2.1 Latent Gaussian Models

Latent Gaussian models (LGMs) fall within a subclass of the structured additive regression models [Rue et al., 2009] meaning the response y_i belongs to the class of exponential families. Hence, the mean $E(y_i) = \mu$ is linked to a structured additive predictor η through a link function $h(\mu)$ such that $h(\mu) = \eta$. For the structured additive regression models η is defined as follows [Fahrmeir et al., 2007],

$$h(\mu) = \eta = \alpha + \sum_{k=1}^{n_\beta} \beta_k z_k + \sum_{j=1}^{n_f} f^{(j)}(u_j) + \epsilon.$$

Here $\{\beta_k\}$ represents the linear effects of explanatory variables $\boldsymbol{z}$, $\{f^{(j)}(.)\}$ represents unknown functions of explanatory variables $\boldsymbol{u}$, and ϵ is an unstructured term. To belong to the class of LGMs the prior distributions of α, $\{\beta_k\}$, $\{f^{(j)}(.)\}$ and ϵ must be Gaussian. All models used in this work belong to the class of LGMs.

2.2 Missing Data

Let $\boldsymbol{y_i}$ be the set of j measurements on the ith subject. Then $\boldsymbol{y_i}$ can be divided into an observed part $\boldsymbol{y_{io}}$ and a missing part $\boldsymbol{y_{im}}$, $\boldsymbol{y_i} = (\boldsymbol{y_{io}}, \boldsymbol{y_{im}})$. Let $\boldsymbol{m_i}$ be the vector of,

$$m_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is missing} \\ 0 & \text{otherwise.} \end{cases}$$

Then the full conditional of $\boldsymbol{y_i}$ and $\boldsymbol{m_i}$ is given as follows,

$$g(\boldsymbol{y_{io}}, \boldsymbol{y_{im}}, \boldsymbol{m_i} | \boldsymbol{\theta}, \boldsymbol{\psi}) \tag{1}$$

where the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ describes the measurement process and missing process, respectively [Gad and Darwish, 2013, Little and Rubin, 2019, Chap. 6.2]. The data is MCAR if the missing process $g(\boldsymbol{m_i}|\boldsymbol{y_{io}}, \boldsymbol{y_{im}}, \boldsymbol{\psi}) = g(\boldsymbol{m_i}|\boldsymbol{\psi})$. The data is MAR if $g(\boldsymbol{m}|\boldsymbol{y_{io}}, \boldsymbol{y_{im}}, \boldsymbol{\psi}) = g(\boldsymbol{m_i}|\boldsymbol{y_{io}}, \boldsymbol{\psi})$. If the data is neither MCAR nor MAR, the data is, by definition, MNAR. If data is MNAR, the missing process must be modeled simultaneously with the measurement process to obtain unbiased inference [Little and Rubin, 2019, Chap. 6.2], and several models are proposed including pattern mixture models and selection models [Little and Rubin, 2019]. In this work a class of selection models known as shared

parameter models (SPMs) is used. From now on, let $\boldsymbol{x_i}$ be the set of fully observed explanatory variables and ϵ_i be an unobserved within-subject random effect with hyperparameter γ. [Little and Rubin, 2019, Chap. 15.2] defined SPM as follows:

$$g(\boldsymbol{y_i}, \boldsymbol{m_i}, \epsilon_i | \boldsymbol{x_i}, \boldsymbol{\theta}, \boldsymbol{\psi}, \gamma) = g(\boldsymbol{y_i} | \boldsymbol{x_i}, \epsilon, \boldsymbol{\theta}) g(\boldsymbol{m_i} | \boldsymbol{x_i}, \epsilon_i, \boldsymbol{\psi}) g(\epsilon_i | \boldsymbol{x_i}, \gamma). \tag{2}$$

This model assumes that both the measurement and dropout processes depend on a shared latent variable ϵ_i. MAR is then a special case with $g(m_i | \boldsymbol{x_i}, \epsilon_i, \boldsymbol{\psi}) = g(m_i | \boldsymbol{x_i}, \boldsymbol{\psi})$ [Vonesh et al., 2006].

3 Case Study: A Blood Pressure Predictive Model based on the HUNT Study.

3.1 The HUNT Study and explanatory analyses

The HUNT Study is a longitudinal population-based health survey in central Norway and the study protocols have been described in detail previously [Krokstad et al., 2013, Åsvold et al., 2021] (Appendix A). Every adult citizen in the now former county of Nord-Trøndelag were invited to participate in clinical examinations and questionnaires in 1984-86 (HUNT1), 1995-97 (HUNT2), 2006-08 (HUNT3), and 2017-19 (HUNT4) [Krokstad et al., 2013, Åsvold et al., 2021]. In this study observations of systolic blood pressure (BP), age (age), body mass index (BMI) and sex (sex, 0 for females and 1 for males) are used. Following Tobin et al. [2005] BP is adjusted by adding 15 mmHg for all participants who self-reported using BP medication. When needed a subscript indicates the HUNT survey of the observation (e.g. BP_2 denotes BP observed at HUNT2). We define a training cohort (HUNT2 cohort) with observations of initial blood pressure ($BP_I = BP_2$), age, BMI and sex from HUNT2, together with future blood pressure ($BP_F = BP_3$) from HUNT3 and a missing indicator m (1 if BP_F is missing in HUNT3, 0 if present). Of 60385 participants in HUNT2, 43.1% of the cohort were missing in HUNT3.

Units and summary statistics for the observations in the HUNT2 cohort are given in Table 1, together with group mean for observations grouped on missing status. In all following analyses BP_F, BP_I, age and BMI observations are standardized by the corresponding sample mean and standard deviation in the HUNT2 cohort.

In Figure 1 and Table 1 we find clear differences between the present and missing participants for age and BP_I (BP_2). Middle-aged participants are less likely to drop out than young or elderly participants, and those with higher blood pressure are more likely to be missing. This suggests that the data is at least MAR. A validation cohort, the HUNT3 cohort, is also constructed. It consists of participants with observations of blood pressure ($BP_I = BP_3$), age, BMI and sex in HUNT3, and future blood pressure ($BP_F = BP_4$) and missing status from HUNT4. Of 50201 participants in HUNT3, 33.3% of the cohort drop out before HUNT4. See Table 5 in Appendix B for further summery of the HUNT3 cohort.

Table 1: The sample mean and standard deviation of BP_F, BP_I, age, and BMI and proportion of female/male participants in the HUNT2 cohort are displayed in the third column. The fourth and fifth columns display sample mean for the present and missing participants in addition to proportions of present/missing participants for the whole cohort and per sex.

Summary of the HUNT2 cohort

Variable	Unit	HUNT2	Present in HUNT3	Missing in HUNT3
BP_3 (BP_F)	mmHg	-	136.1	-
BP_2 (BP_I)	mmHg	139.5 (23.6)	135.2	145.0
age_2	years	50.0 (17.1)	47.0	54.01
BMI_2	kg/m^2	26.4 (4.1)	26.2	26.6
sex			56.9 %	43.1 %
female	0	53.0 %	59.3 %	40.7 %
male	1	47.0 %	54.3 %	45.7 %

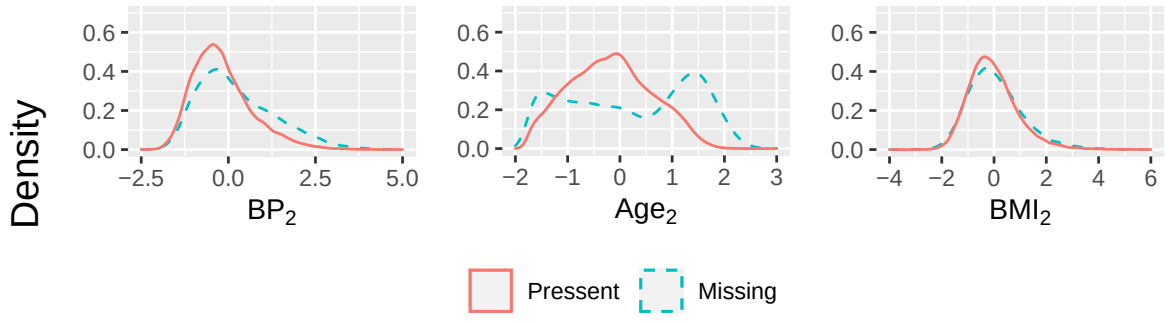


Figure 1: Smoothed empirical density of BP_I (BP_2), age, and BMI for all participants in the HUNT2 cohort.

3.2 A Shared Parameter Model for Blood Pressure

We set up a shared parameter model (SPM) for BP_F and the missing process using age, sex, BMI, and BP_I, as explanatory variables in the framework of a LGM as presented in Section 2.1. Let BP_{Fi} and m_i represent future blood pressure and missing status for individual i. The likelihoods are chosen to be Gaussian with identity link for BP_F and Bernoulli with logit link for m_i; $BP_{F_i} \sim N(\eta_{BP_i}, \sigma^2_{BP})$ and $m_i \sim Bernoulli(p_i)$ with $logit(p_i) = \eta_{m_i}$. From the general formula of LGMs in Section 2.1 the explanatory variables can be included either as linear effects or as non-linear effects. Based on the work by Espeland [2020] and the analyses in Appendix C we chose to include all explanatory variables but age in the missing process as linear effects. Further we introduce a shared parameter ϵ_i in both linear predictors, and with an association parameter c for the missingness model;

$$\begin{aligned} \eta_{BPi} &= \alpha_0 + \alpha_{BP} BP_{Ii} + \alpha_{age} age_i + \alpha_{BMI} BMI_i + \alpha_{sex} sex_i + \epsilon_i \\ \eta_{mi} &= \beta_0 + \beta_{BP} BP_{Ii} + f(age_i) + \beta_{BMI} BMI_i + \beta_{sex} sex_i + c\epsilon_i, \end{aligned} \tag{3}$$

where the shared parameters ϵ_i are assumed to be independent Gaussian $\epsilon_i \sim N(0, \sigma^2_\epsilon)$ and $f()$ is a random walk of order two with variance σ_{age}, as defined in Appendix C. To avoid identifiability issues between the shared parameter ϵ and the likelihood of BF_F we fix σ^2_{BP} to a small value ($\sigma^2_{BP} = 0.001^2$). All regression parameters α_0, α_{BP}, α_{age}, α_{BMI}, α_{sex}, β_0, β_{BP}, β_{BMI}, and β_{sex} are given independent priors $N(0, 10^{3^2})$, and σ^2_{age} and σ^2_ϵ are assigned independent gamma priors, Gamma$(1, 5 \cdot 10^5)$. We expect the shared parameter to influence the missing process similarly or less then the standardized explanatory variables. Therefore, the association parameter c is given an informative prior $c \sim N(0, 1^2)$. A sensitivity study is conducted for this prior, see Appendix E.

When the association parameter $c = 0$, the models for BP_F and m are independent and we have a model that assumes data MAR. We refer to this model as the naive model, and it is used as a benchmark model.

For simplicity we introduce some notation. Let $\boldsymbol{x_i} = (BP_{Ii}, age_i, BMI_i, sex_i)$ be the explanatory variables and $\boldsymbol{y_i} = (BP_{Fi}, m_i)$ the response variables for individual i. Further let $X = (x_i, .., x_n)^T$ be the explanatory variables for all n participants and $Y = (y_1, ..., y_n)^T$ be the corresponding response variables in a cohort. When needed we use superscript to indicate the HUNT2 or HUNT3 cohort, i.e., X^2 are the explanatory variables from the HUNT2 cohort. Denote the modeling parameters by $\boldsymbol{\theta} = (\alpha_0, \alpha_{BP}, \alpha_{age}, \alpha_{BMI}, \alpha_{sex}, \beta_0, \beta_{BP}, \beta_{BMI}, \beta_{sex}, \epsilon, f, c, \sigma_\epsilon, \sigma_{age})$ where f refer to the Gaussian variables of the additive effect for age.

3.3 Inference

Conditioned on data (X, Y) we can achieve posterior distributions for the parameters, $\pi(\boldsymbol{\theta}|X, Y)$. In this work, we are either interested in the marginal posterior of selected parameters (j), $\pi(\theta^{(j)}|X, Y)$, or in the posterior predictive distribution for a new person with explanatory variables $\boldsymbol{x_{new}}$. This posterior predictive distribution is given by

$$\begin{aligned} \pi(\boldsymbol{y_{new}}|\boldsymbol{x_{new}}, X, Y) &= \int \pi(\boldsymbol{y_{new}}, \boldsymbol{\theta}|\boldsymbol{x_{new}}, X, Y) d\boldsymbol{\theta} \\ &= \int \pi(\boldsymbol{y_{new}}|\boldsymbol{x_{new}}, \boldsymbol{\theta}) \pi(\boldsymbol{\theta}|X, Y) d\boldsymbol{\theta}. \end{aligned}$$

The SPM suggested in Section 3.2 is a LGM, as described in Section 2.1 that meets the requirements for using the computationally efficient integrated nested Laplace approximations (INLA), see Steinsland et al. [2014] for more details for an analogous SPM. The latent Gaussian field consists of (α_0, α_{BP}, α_{age}, α_{BMI}, α_{sex}, β_0, β_{BP}, β_{BMI}, β_{sex}, ϵ, f, c) and the non-Gaussian hyperparameters are (σ_ϵ, σ_{age}).

3.4 Validation Scheme Using the HUNT3 Cohort

We evaluate the prediction models obtained from the HUNT2 cohort using the HUNT3 cohort. For each participant i in the HUNT3 cohort we get the predictive distributions $\hat{\boldsymbol{y}}_{\boldsymbol{i}} \sim \pi(\boldsymbol{y_i^3}|\boldsymbol{x_i^3}, X^2, Y^2)$ and specifically for the future blood pressure and missing status $\hat{BP}_{Fi} \sim \pi(BP_{Fi}^3|\boldsymbol{x_i^3}, X^2, Y^2)$ and $\hat{m}_i \sim \pi(m_i^3|\boldsymbol{x_i^3}, X^2, Y^2)$.

To evaluate the predictive performance we calculate the mean continuous rank probability score (CRPS) of $\hat{BP}_{Fi}$ and mean Brier score of $\hat{m}_i$ over all participants in the HUNT3 cohort. Let $F_{\hat{BP}_{Fi}}(x)$ be the cumulative probability distribution of $\hat{BP}_{Fi}$ and BP_{Fi} the observed blood pressure in HUNT4 for participant i, then $CRPS(F, y) = \int_{-\infty}^{\infty}[F(x) - H(x - BP_{Fi})]^2 dx$ where $H(u)$ is the Heaviside function (0 for $u < 0$ and 1 for $u > 0$). The blood pressure model can only be validated on the participants observed in both the HUNT3 and the HUNT4 surveys. In contrast, missing model can be evaluated for all participants.

Predictions from the SPM and the naive model are compared by their posterior mean for the HUNT3 cohort participants as well as their CRPS and Brier scores.

3.5 Evaluation of Missing not at Random by Conditioning on Missing Status

We introduce a novel method for validating if data is MNAR based on a SPM fitted to a training dataset (here the HUNT2 cohort) and the difference in predictive performance for a validation dataset (here the HUNT3 cohort) for predictors with and without conditioning on the missing status in the training dataset. For readability we introduce the method using the notation of HUNT2 cohort and HUNT3 cohort, but the method is general.

If data are MNAR and the SPM is true, there is information about the shared parameter in the missing status, and conditioning on the missing status, i.e. the value of m_{new} should give a better predictor. For each participant i in the HUNT3 cohort we can from the the predictive distribution $\pi(\boldsymbol{y_i^3}|\boldsymbol{x_i^3}, X^2, Y^2)$ derive both the marginal predictive distribution for the future blood pressure $\hat{BP}_{Fi} \sim \pi(BP_{Fi}^3|\boldsymbol{x_i^3}, X^2, Y^2)$ and the predictive distribution for the future blood pressure conditioned on the missing status $\hat{BP}_{Fi}|m_i \sim \pi(BP_{Fi}|\boldsymbol{x_i^3}, X^2, Y^2, m_i^3)$.

In practice, we can for a validation dataset only evaluate the predictions for the present participants, and not the dropouts, and we therefore compare the predictive performance of $\hat{BP}_{F_{new}}$ and $\hat{BP}_{F_{new}}|m_{new} = 0$ for all presents participants. In this work we have calculated the absolute error of the posterior mean predictions for each participant, and compare mean absolute errors (MAE) for the prediction with and without conditioning on missing status.

3.6 Software and Code

In this work, we use the R-INLA software [R-INLA, 2021]. The R-INLA software supports fitting models with multiple likelihoods [Steinsland et al., 2014, Espeland, 2020, Gómez-Rubio, 2020, Chap. 6.4] which is the case for the SPM (3). All the code is available at the GitHub repository by Hofman [2021]. Since data can not be shared, to protect participants privacy, also a compleatly simulated dataset is provided.

4 Results for the Blood Pressure Case Study

The shared parameter models (SPM) and the naive model introduced in Section 3 are fitted using the HUNT2 cohort as described in Section 3.2. This chapter presents and compares the posterior distributions of interest. Further, the predictive models are evaluated through predictive performance of the HUNT3 cohort, as described in Section 3.4 and Section 3.5.

4.1 Results for the HUNT2 Cohort

The posterior distributions of the estimates obtained by the SPM and the naive model, introduced in Section 3.2 and fitted to the HUNT2 cohort, can be seen in Figure 2. The posterior mean and 95% credible intervals are presented in Table 6 in Appendix D.

For the blood pressure submodel in the SPM, we see that the effect of BP_I (α_{BP}) is the largest followed by age, BMI and sex. The effect of BMI and sex are close to zero. The parameter estimates of the naive blood pressure model are of the same order as the SPM. However, in the blood pressure submodel all variables but age have weaker effects in the naive model than for the SPM. The difference is especially pronounced for α_0 and α_{BP}, suggesting the two models could result in different predictions.

For the missing process in the SPM, we see that sex has the largest effect on the probability of dropping out, followed by BP_I and BMI. The age effect is largest for the elderly and smallest for middle-aged participants (Figure 3). The SPM and the naive model are more similar in parameter estimates for the missing process than the blood pressure process. However, the parameter estimates of the naive model are shifted towards lower values than the SPM.

The association parameter c, connecting the two submodels (3) in the SPM is clearly positive, which implies an increase in the probability of dropping out for larger random effect. According to the SPM, this means that participants with higher BP_F than can be explained by the explanatory variables are more likely to drop out.

4.2 Results for Toy Example Participants

Since the data used in this work contains personal information, we consider three constructive toy example participants to explore the model presented in Section 3.2 and Section 3.4 on an individual level. We used a young and underweight female with low BP_I (id1), a middle-aged and overweight female with average BP_I (id2), and an old and obese female with severely high BP_I (id3), see Table 2. For these participants, the effect of the association parameter on the probability of dropping out is plotted in Figure 4 for random effects between -1.5 and 1.5 which corresponds to approximately two standard deviations of the

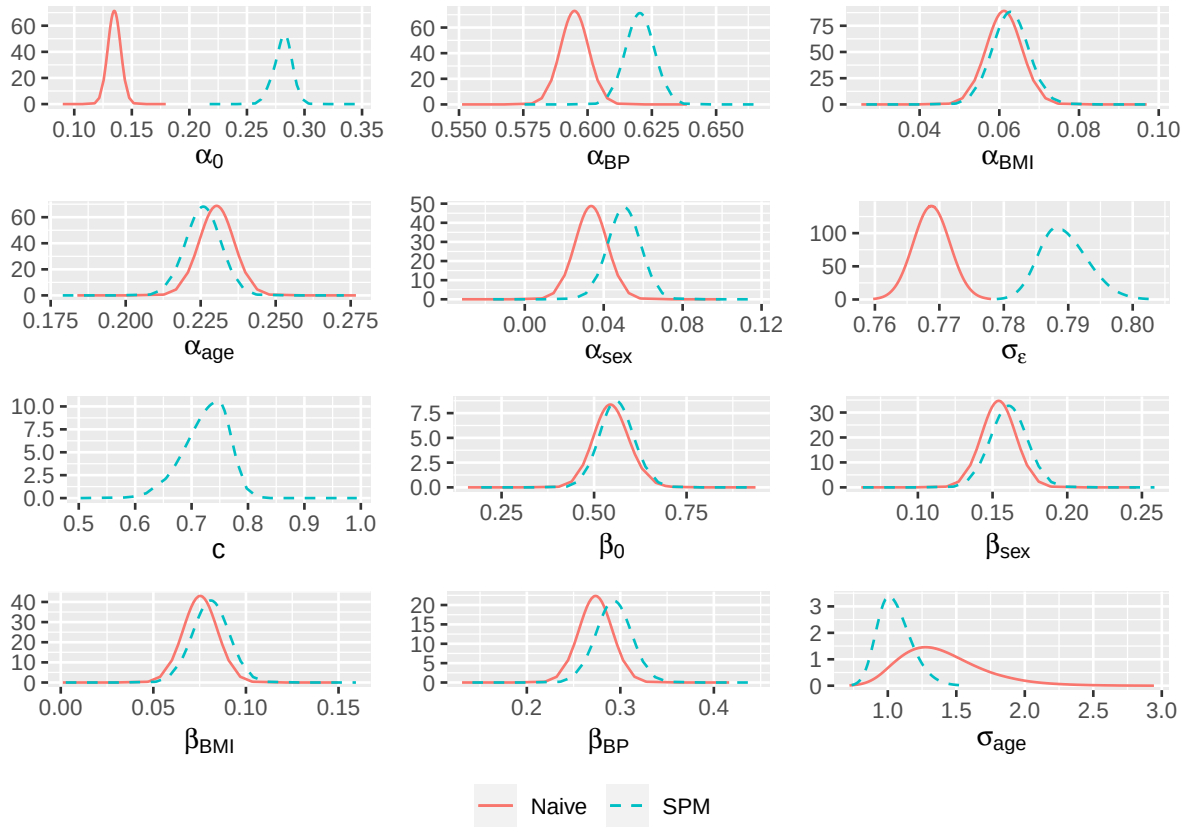


Figure 2: Posterior distribution of the latent field and hyperparameters for the SPM and naive model fitted to the HUNT2 cohort.

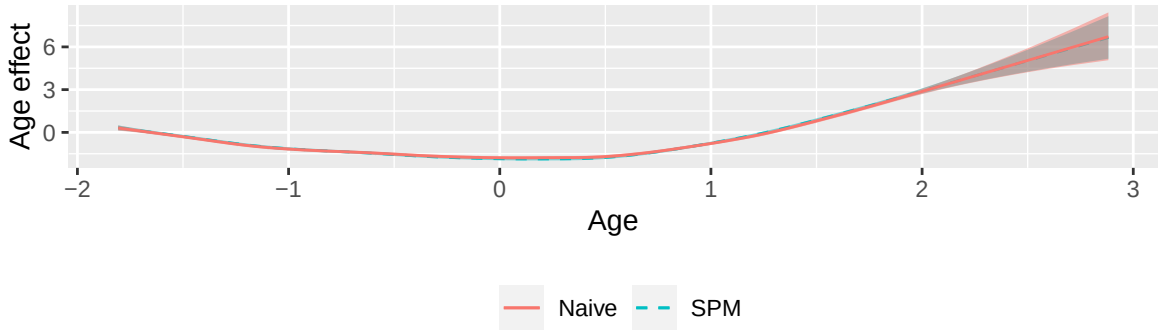


Figure 3: Age effect for the SPM and naive model fitted to the HUNT2 cohort with 95% credible bands.

Table 2: Values of BP_I, age and BMI for three female toy example participants.

id	BP_I^*	age^*	BMI^*	sex	BP_{Itrue}	age_{true}	BMI_{true}
1	-2	-1.5	-2	female	92.2	24.4	18.2
2	0	0	0	female	139.5	50.0	26.4
3	2	1.5	2	female	186.7	75.7	34.6

* Standardized values

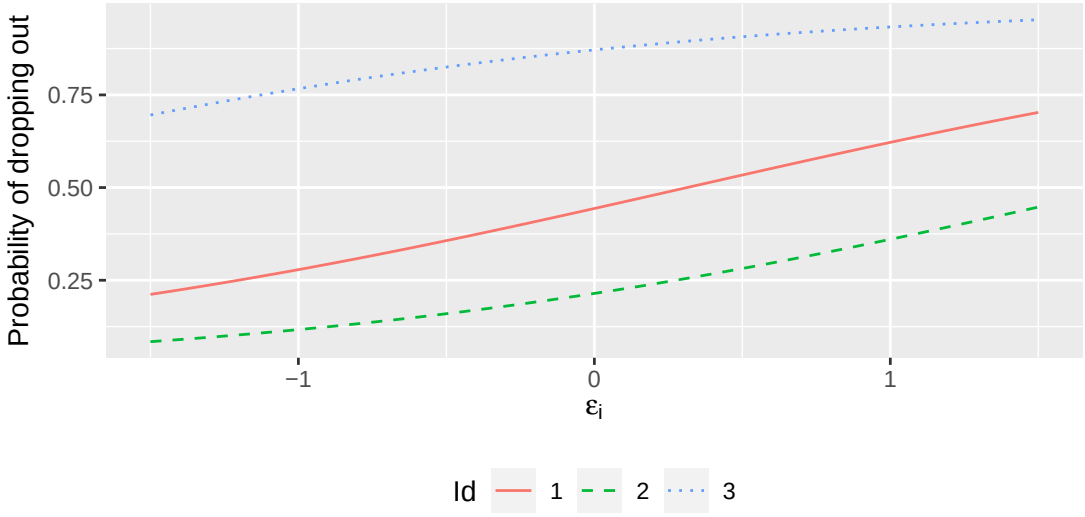


Figure 4: The probability of dropping out as a function of the individual random effect of the blood pressure model for three toy example participants specified in Table 2.

random effect. We find that a larger value of the random effect gives a large probability of dropping out for all three toy example participants.

The posterior predictive distributions of BP_F for the three simulated participants are plotted in Figure 5 for both the SPM and the naive model. For all toy example participants the posterior predictive distribution from the SPM is shifted towards larger values than the naive model, and more so for id1 and id3 who have more extreme explanatory variables.

4.3 Validation of Model Predictions for the HUNT3 Cohort

Table 3: CRPS score for predictions of systolic blood pressure in HUNT4 for the present participants. Brier score for mean predictions of the probability of drop out in HUNT4. All predictions are based on the HUNT3 cohort. The best score is indicated in bold.

	CRPS	**Brier**		
		All	Present	Missing
SPM	0.4406	0.2082	0.1656	**0.2937**
Naive	**0.4337**	**0.2072**	**0.1602**	0.3014

For all HUNT3 participants predictive distributions are calculated as described in Section 3.3 estimated from the HUNT2 cohort for both the naive model and the SPM. The

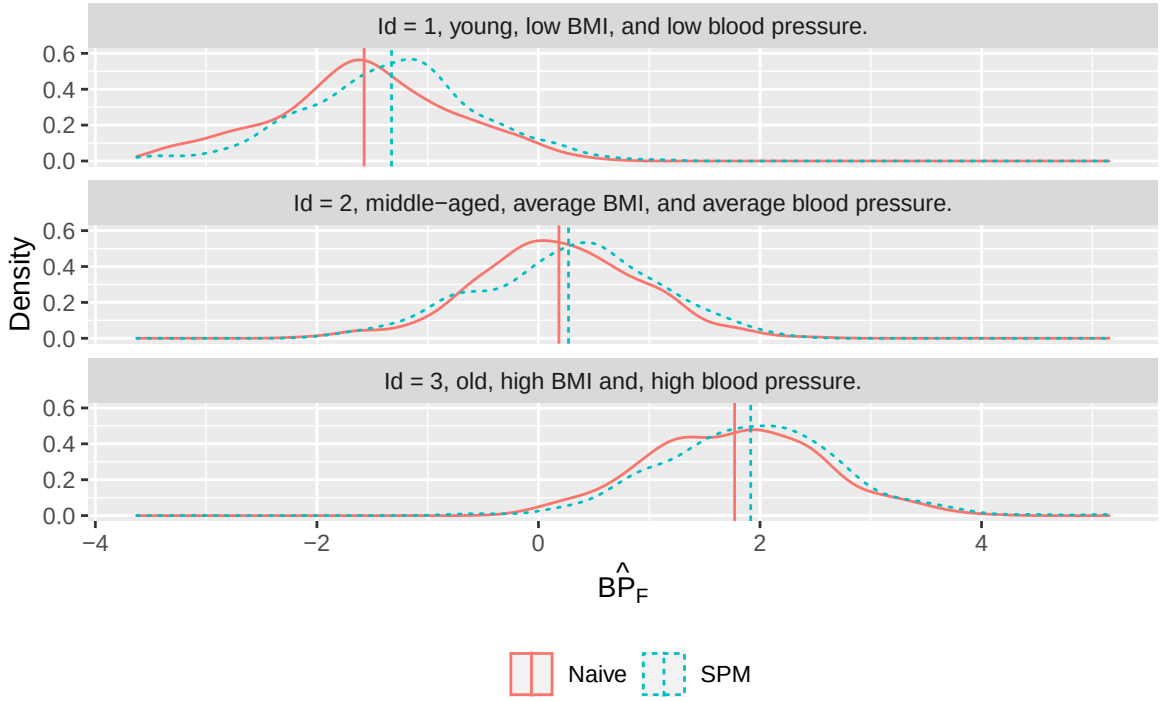
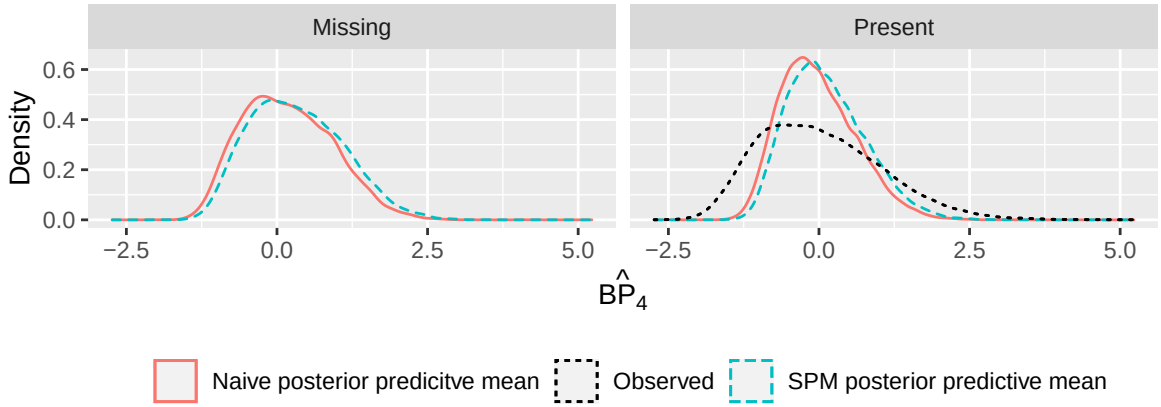
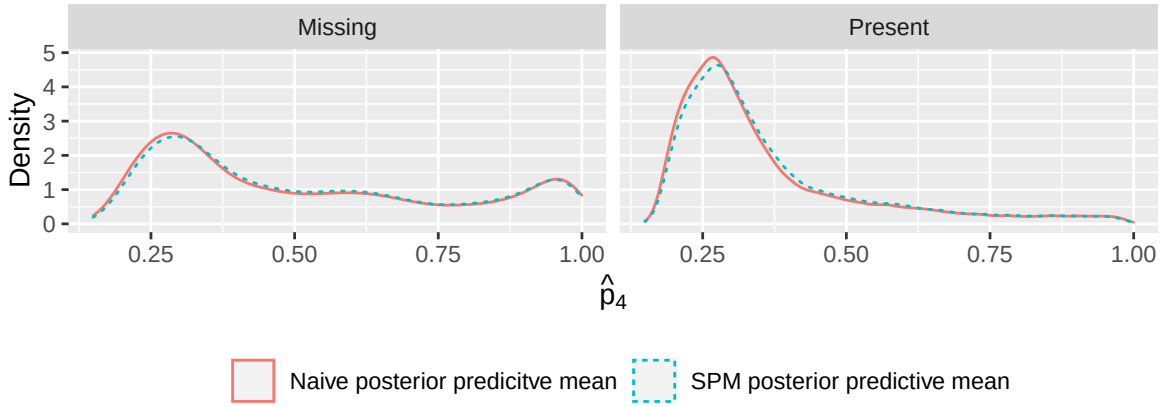


Figure 5: Posterior predictive distribution of the future blood pressure (BP_F) for three toy example participants in Table 2 for the naive model and the SPM models. Vertical lines indicate posterior means.



(a) Density over all participants in HUNT3 of the posterior mean of predicted future blood pressure $\hat{BP}_{Fi}$.



(b) Density over all participants in HUNT3 of the posterior mean of predicted probability of dropping out $\hat{p}_i$.

Figure 6: All plots display the participants grouped by missing status in the HUNT3 cohort.

empirical distributions of posterior mean predictions for participants in HUNT3 cohort are found in Figure 6a, and for the BP_F and in Figure 6b for probability of drop out p. For reference the empirical distribution for observed BP_F is included for present participants in Figure 6a. We see that on a population level the SPM predicts slightly larger values for the BP_F than the naive model. The distributions of mean predictions of the probability of dropping out are very similar. The SPM predicts a slightly higher probability of dropping out, on average, than the naive model. We note that the differences between naive and SPM have little practical implication on population level, but for some individuals there can be differences of some practical significance.

We further evaluate the predictive performance for the HUNT3 cohort for the SPM and the naive model from the HUNT2 cohort as described in Section 3.4. Mean CRPS for BP_F and mean Brier scores for the probability of drop out are given in Table 3. The mean CRPS is very similar for the SPM and naive model, but the CRPS for the naive model predictions is slightly smaller and hence the naive model performs slightly better. This might be explained by the fact that the missing participants do not affect the likelihood of

Table 4: Mean absolute error for $\hat{\boldsymbol{BP}}_{\boldsymbol{F}}|m$ and $\hat{\boldsymbol{BP}}_{\boldsymbol{F}}$. The best score is indicated in bold.

	$\hat{\boldsymbol{BP}}_{\boldsymbol{F}}\|m$	$\hat{\boldsymbol{BP}}_{\boldsymbol{F}}$	MAE($\hat{\boldsymbol{BP}}_{\boldsymbol{F}}\|m$) - MAE($\hat{\boldsymbol{BP}}_{\boldsymbol{F}}$)
MAE	**0.6140**	0.6155	-0.0014

the naive model. Hence, the model is optimized to perform well for the present participants. We explore this further in a simulation study in Section 5.2.

The Brier scores in Table 3 are slightly better for the naive model then for the SPM when evaluating for all participants. However, when grouped by missing status, the naive model performs better on the present participants, and the SPM performs better on the dropouts.

4.4 Evaluate of the Missing Not at Random Assumption

We evaluate the MNAR assumption by comparing the predictive performance of the SPM with and without conditioning on the missing status for the HUNT3 cohort as described in Section 3.5. The results are given in Table 4. We see that the mean absolute errors (MAEs) are very similar, but that the predictions for $\boldsymbol{BP_4}$ given $\boldsymbol{m_4}$ yield a slightly smaller error than not knowing $\boldsymbol{m_4}$. This suggests that some information from the missing process affects the $\boldsymbol{BP_F}$. A simulation study is conducted in Section 5.2, and a difference of -0.0014 is within what is to be expected when the SPM is true for similar training and validation datasets.

5 Simulation Studies

We set up several simulation studies to explore the properties of the SPM, the naive model and the method validating MNAR by conditioning on missing status for datasets with the size and structure of the HUNT2 and HUNT3 cohort. In all simulation studies data sets for training models are simulated using the same number of participants and explanatory variables as in the HUNT2 cohort, i.e. $X = X^2$. Further, when simulating data, the parameters $\boldsymbol{\theta} = \boldsymbol{\theta^{true}}$ are set to the posterior mean estimates of the SPM and naive model fitted on the HUNT2 cohort, see Table 6, in Appendix D, in all simulation studies. When predictions are studied, simulated validation datasets are based on the same size and explanatory variables as the HUNT3 cohort, i.e. $X = X^3$ is used.

5.1 Simulation Study exploring Bias and Coverage

The aim of this simulation study is to study the properties of the posterior estimates for the SPM and the naive model in a situation similar to the HUNT2 cohort. 100 independent new response data (i.e. both future blood pressure $\boldsymbol{BP_F}$ and the missing status $\boldsymbol{m}$) are simulated using the SPM with parameters $\boldsymbol{\theta^{true}}$ and with explanatory variables as in the HUNT2 cohort, resulting in $Y^{(l)}$ for $l = 1 \dots 100$. For each of the data sets $(X^2, Y^{(l)})$ both the SPM and the naive model are fitted. This gives posterior distributions $\pi(\boldsymbol{\theta}|X^2, Y^{(l)})$. Each of these are summarized by the posterior mean and the coverage indicator of the true value in 95% credibility interval.

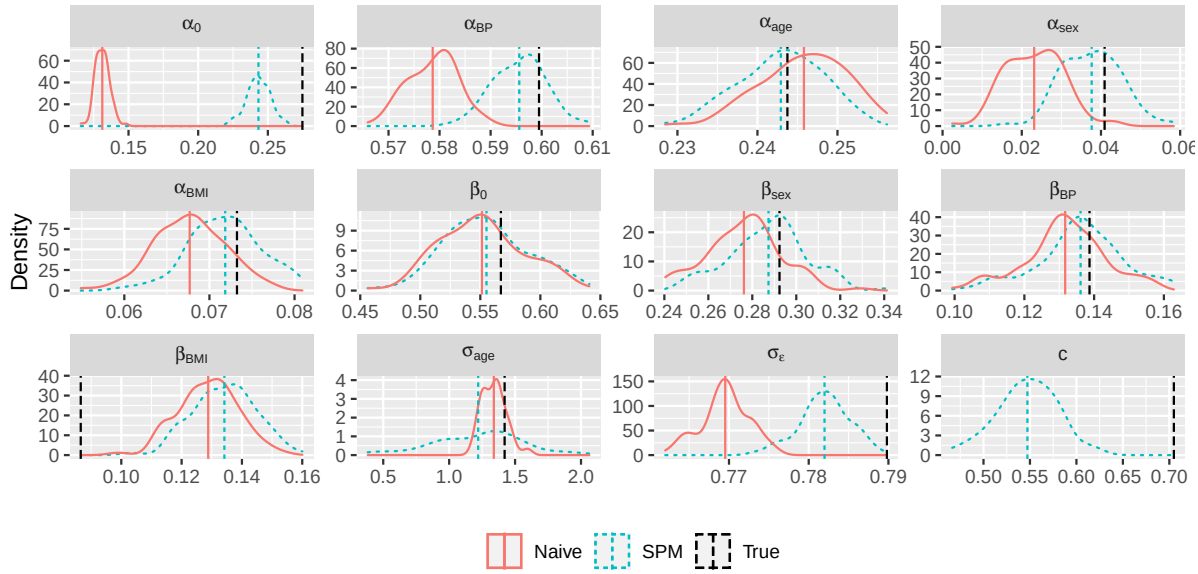


Figure 7: Distribution of posterior mean estimates on simulated data MNAR following the SPM as described in Section 5.1. The mean of posterior means for both the SPM and naive model and the true value are indicated by the vertical lines.

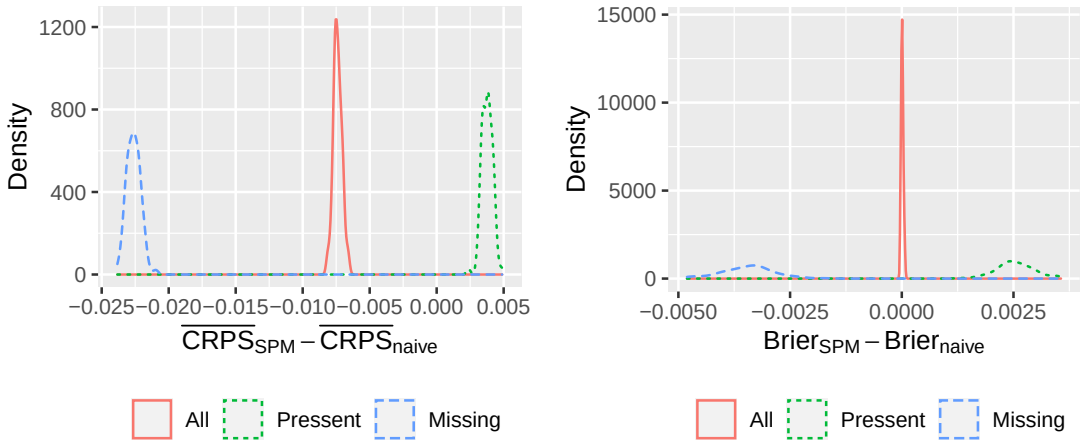
The resulting mean posterior mean, bias, and coverage are found in Table 8 in Appendix F. The distribution of the posterior means from the simulation study when the true parameters are the posterior mean estimates of the SPM can be seen in Figure 7. From Figure 7 we find that both the naive model and SPM are biased when the data is generated from the SPM. However, the SPM is less biased especially for the blood pressure model parameters (i.e. $\alpha_0, \alpha_{PB}, \alpha_{age}, \alpha_{sex}$, and α_{BMI}). The association parameter c has especially low coverage, see Table 8 in Appendix F).

We have also preformed a similar simulation study, but with data MAR by setting the association parameter $c = 0$ when simulating data. The results are presented in Appendix F When the data is MAR, both the SPM and naive model have very little bias, and in particular the association parameter c is centered around zero and has good coverage.

5.2 Simulation Study for Predictions

The aim of this simulation study is to learn about the predictive performance. Following the procedure in Section 5.1 a data sets mimicking the HUNT2 cohort, (X^2, Y), and $l = 1 \dots 100$ data sets mimicking the HUNT3 cohort, $(X^3, Y^{(l)})$, are simulated using the SPM with parameters $\boldsymbol{\theta}^{\boldsymbol{true}}$. Corresponding posterior distributions based on the simulated HUNT2 cohort $\pi_{SPM}(\boldsymbol{\theta}|X^2, Y)$ and $\pi_{naive}(\theta|X^2, Y)$ are found based on the SPM and the naive model, respectively. From these, posterior predictive distributions are achieved, and the predictive performance is evaluated for each participant in the simulated data set $(X^3, Y^{(j)})$ as described in Section 3.4 by the CRPS and Brier score. A large advantage for the simulated data sets is that predictive performance can be evaluated not only for present participants, but also for those that are missing, and we include them.

Further we compare the predictive performance conditioned on missing status



(a) Difference in mean CRPS score for the SPM and naive model with 100 simulations

(b) Difference in mean Brier score for the SPM and naive model with 100 simulations

Figure 8: Plots for differences in mean scores for participants grouped by missing status in the HUNT3 cohort.

$\pi_{SPM}(Y^{(l)}|\boldsymbol{m}^l, X^3, \boldsymbol{\theta})$ with $\pi_{SPM}(Y^{(l)}|X^3, \boldsymbol{\theta})$ through MAE as described in Section 3.5.

Figure Figure 8a displays the distribution of the difference between the mean CRPS for the SPM and the naive model. This difference is displayed for all simulated participants (present/missing) and grouped on missing status. We see that the SPM performs better for all participants and the dropouts, but the naive model performs better on present participants. This demonstrate that for our case study even if the data is MNAR and follows the SPM, the naive model is expected to obtain a better CRPS than the SPM when only evaluating for present participants (and the missing participants can not be used for evaluation!).

Figure Figure 8b shows the distribution of the difference in Brier score for all, missing and present participants. We see similar results here, the SPM predicts best for missing participants and the naive model predicts best for the present participants.

For data MNAR, we can see from Figure 10a that the distribution of MAEs for the 100 simulations is shifted towards lower values when the missing status is known when the data is MNAR. Further, we see from Figure 9 that the MAE was smaller for every simulated dataset when predicting $\boldsymbol{B}\hat{\boldsymbol{P}}_{\boldsymbol{F}}|\boldsymbol{m}$ than $\boldsymbol{B}\hat{\boldsymbol{P}}_{\boldsymbol{F}}$ since zero is not contained in the distribution of $MAE(\boldsymbol{B}\hat{\boldsymbol{P}_{\boldsymbol{F}}}|\boldsymbol{m}) - MAE(\boldsymbol{B}\hat{\boldsymbol{P}}_{\boldsymbol{F}})$. Hence, for data MNAR following the SPM, we can expect the predictions of $\boldsymbol{B}\hat{\boldsymbol{P}_{\boldsymbol{F}}}|\boldsymbol{m}$ to be better than $\boldsymbol{B}\hat{\boldsymbol{P}}_{\boldsymbol{F}}$.

For data MAR Figure 10b shows we obtain almost the same mean absolute error for every simulated dataset when predicting $\boldsymbol{B}\hat{\boldsymbol{P}}_{\boldsymbol{F}}|\boldsymbol{m}$ as when predicting $\boldsymbol{B}\hat{\boldsymbol{P}}_{\boldsymbol{F}}$. In addition, we see from Figure 9 that the distribution of $MAE(\boldsymbol{B}\hat{\boldsymbol{P}}_{\boldsymbol{F}}|\boldsymbol{m}) - MAE(\boldsymbol{B}\hat{\boldsymbol{P}}_{\boldsymbol{F}})$ covers zero.

From Figure 9, we see that there is no overlap between the two distributions for, $MAE(\boldsymbol{B}\hat{\boldsymbol{P}_{\boldsymbol{F}}}|\boldsymbol{m}) - MAE(\boldsymbol{B}\hat{\boldsymbol{P}}_{\boldsymbol{F}})$. Also, the distributions are narrow both when the data is MNAR and MAR. This demonstrates that even minor differences in MAE between predictions of $\boldsymbol{B}\hat{\boldsymbol{P}}_{\boldsymbol{F}}|\boldsymbol{m}$ and $\boldsymbol{B}\hat{\boldsymbol{P}}_{\boldsymbol{F}}$ indicate that data are MNAR.

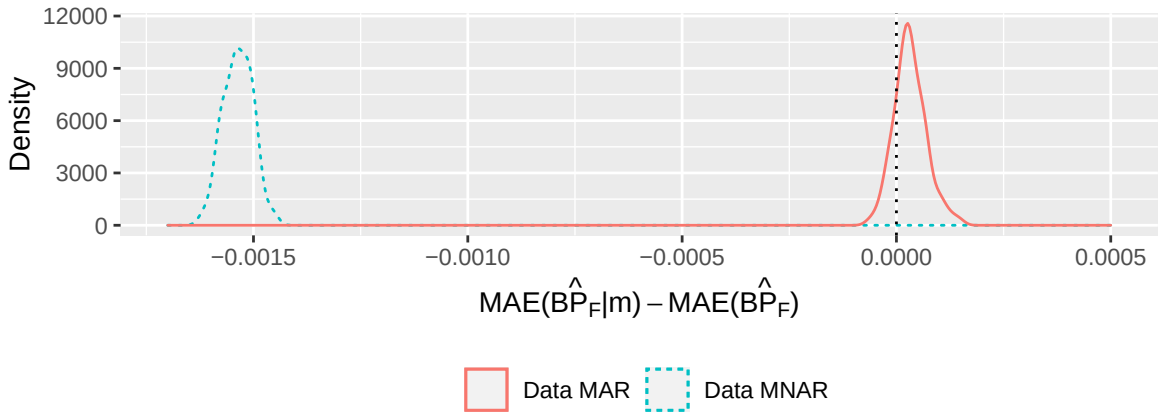
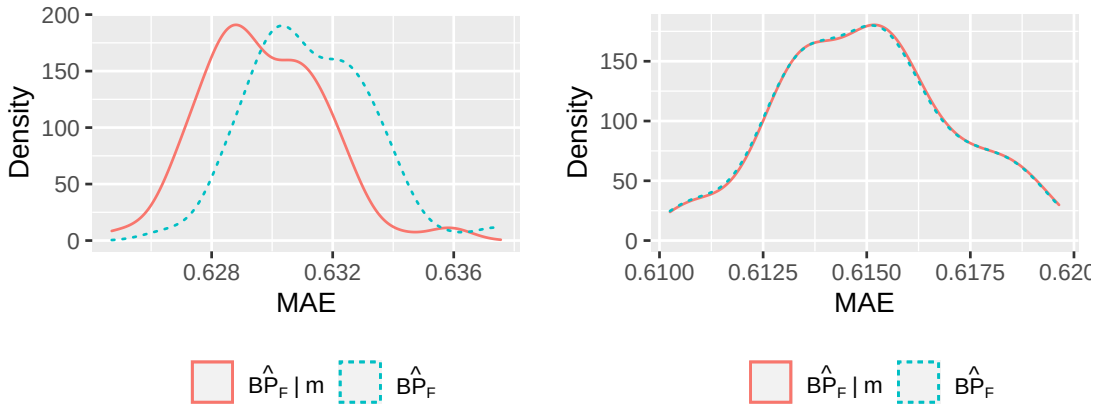


Figure 9: Difference in mean absolute error (MAE) of the predictions of $\boldsymbol{B\hat{P}_F|m}$ and $\boldsymbol{B\hat{P}_F}$ when the data is MNAR and MAR. The vertical line indicates zero.



(a) Data MNAR

(b) Data MAR

Figure 10: Distribution of the mean absolute error MAE of the posterior mean prediction for future blood pressure obtained from the validation scheme presented in Section 5.2

6 Discussion

In this work, we propose reasonable models in the form of a SPM for predicting future blood pressure (BP_F) and missing status. The SPM is compared to a naive model which assumes data MAR. Our findings indicate that blood pressure missing due to dropout in the HUNT Study is MNAR. The simulation study also confirms that if the underlying assumption of data MNAR is valid, the SPM accounts for this better than the naive model, even though both models are biased. We note that there seem to be some identifiability issues resulting in a bias when the data is MNAR and a low coverage for some model parameters in the simulation studies (Figure 7, Figure 14, Table 8, and Table 9). This bias is especially pronounced for the association effect c when the data is MNAR. Our initial thought was that the low coverage could be due to misspecification of the informative prior used for c. However, the prior sensitivity study presented in Appendix E indicates that the model is robust to prior specifications concerning the association parameter.

We performed all simulation studies related to bias and coverage on data sets that mimic our study system in size and explanatory variables, i.e. the HUNT2 cohort for training and the HUNT3 cohort for validation, We have not explored asymptotic properties of the models nor how uncertainty and biases change with the size of the datasets.

The new validation scheme proposed in Section 3.5 to evaluate the MNAR assumption indicates that the missing status contains information about BP_F given data following the SPM. According to the simulation studies on this validation scheme (Section 5.2), the model predictions conditioned on missing status are only better than those without knowledge of missing status if the data is MNAR (Figure 9). When the data is MAR, the mean absolute error of the two posterior mean predictions are almost identical. Hence we have strong indications that blood pressure missing in the HUNT Study is MNAR.

Weather data is MAR or MNAR relies on untestable assumptions [Enders, 2011]. However, for future research it would be interesting to study the models and results of this work in the framework of missingness graphs introduced by Mohan and Pearl [2021].

As a final note, we acknowledge that the models presented in this work are far from trivial to use and are computationally demanding. Especially the use of the software package R-INLA can be challenging for non-statisticians. However, we find that the models presented provide helpful insight into the underlying structure of the HUNT data and how one can model blood pressure accounting for data MNAR.

To summarize this work, we find that the SPM is better suited to account for data MNAR than the naive model even though both models are biased. We have shown that the SPM can be fitted efficiently through INLA and can be used for predictive purposes. In addition, we have proposed a novel scheme for evaluating if data is MNAR if the data follows the SPM. This scheme is tested in simulation studies and indicates that the blood pressure missing in the HUNT Study is MNAR.

7 Acknowledgement

The Trøndelag Health (HUNT) Study is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology NTNU), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. Participation in the HUNT Study is voluntary, and all participants provided written informed consent before participation. The Regional Committee on Medical and Health Research Ethics of Norway (REK; 2018/1824) approved this work in July 2021.

A The Trøndelag Health Study Protocol

The HUNT Study protocols are described in detail in Krokstad et al. [2013] and Åsvold et al. [2021]. Here, we briefly overview the performed data collection relevant to this work. Age and sex were extracted from the Norwegian Population Registry. Height and weight were measured after removing shoes and other heavy clothing. The BP was measured, by trained personnel, in a sitting position after two minutes of rest. Three measurements were taken, one minute apart, of which the mean of the second and third were used to report BP. To assess the current use of BP medication, self-reported questionnaires were used. In HUNT2, the current use of BP medication was captured in "Are you taking medication for high blood pressure?" [Never; Previously; Currently]. In HUNT3, the question was reformulated to "Do you take, or have you taken medication for high blood pressure?" [No; Yes]. Therefore, we combine this question with the answer to "If you are currently taking medicine for high blood pressure, have you felt unwell/ had side effects from this medicine?". We assume only the participant who currently takes BP medicine answered this question. In HUNT4, the use of BP medicine was captured through the question "Do you currently use any prescription medication for high blood pressure?" [No; Yes].

B Summary of the HUNT3 cohort

The validation cohort (HUNT3 cohort) consists of participants with observations of BP, age, BMI and sex in HUNT3, and BP and missing status in HUNT4. Of 50201 in HUNT3, 33.3% drop out prior to HUNT4. Summary statistics for the HUNT3 cohort are given in Table 5, together with group mean for HUNT2 observations grouped on missing status.

Table 5: The sample mean and standard deviation of BP_F, BP_I, age, and BMI and proportion of female/male participants in the HUNT3 cohort are displayed in the third column. The fourth and fifth columns displays sample mean for the present and missing participants in addition to proportions of present/missing participants for the whole cohort and per sex.

Summary of the HUNT3 cohort

Variable	Unit	HUNT3	Present in HUNT4	Missing in HUNT4,
BP_4	mmHg	-	136.48	-
BP_3	mmHg	133.21 (20.71)	131.48	136.67
age_3	years	53.08 (16.01)	51.68	55.90
BMI_3	kg/m^2	27.17 (4.41)	27.12	27.28
sex			66.7 %	33.3 %
female	0	54.6 %	68.7 %	31.3 %
male	1	45.4 %	64.3 %	35.7 %

C Shared Parameter Model With Additive Effects

Based on the work done by Espeland [2020] we explore the need for non-linear effects in the SPM introduced in Section 3.2. We model all continuous variables as additive effects ($f(z)$) through a random walk of order 2 with a sum to zero constraints [Gómez-Rubio, 2020]. Let $m = 1 : n$ be the index for the increments and define $\Delta^2 z_m = z_m - 2z_{m+1} + z_{m+2} \sim N(0, \sigma^2)$. The density for $f(\boldsymbol{z})$ is, $f(\boldsymbol{z}|\sigma) \propto \sigma^{-\frac{n-2}{2}} exp\left\{ -\frac{1}{2\sigma} \sum_{m=1}^{n-2} (\Delta^2 z_m) \right\}$ Further, the sum of all random effect components is constrained to be zero. For more information about random walk priors and sum to zero constraints, see Rue and Held [2005].

The specification of linear predictors of the SPM defined in Section 3.2 becomes,

$$\begin{aligned} \eta_{BPi} &= \alpha_0 + f_{BP_F}(BP_{Ii}) + f_{BP_F}(age_i) + f_{BP_F}(BMI_i) + \alpha_{sex} sex_i + \epsilon_i \\ \eta_{mi} &= \beta_0 + f_m(BP_{Ii}) + f_m(age_i) + f_m(BMI_i) + \beta_{sex} sex_i + c\epsilon_i. \end{aligned} \tag{4}$$

All regression parameters α_0, α_{sex}, β_0, and β_{sex} are given independent priors $N(0, 10^{3^2})$. The shared parameters ϵ_i are assumed to be independent Gaussian $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. Both the additive effects $f(.)$ and ϵ_i have hyperparameters ($\sigma_{BP_F BP_I}$, $\sigma_{BP_F age}$, $\sigma_{BP_F BMI}$, σ_{mBP_I}, σ_{mage}, σ_{mBMI}, σ_ϵ) with independent gamma priors, Gamma$(1, 5 \cdot 10^5)$. The association parameter c is given an informative prior $c \sim N(0, 1^2)$.

Modeling all parameters in an additive way is extremely computationally demanding. With 32 CPU cores and 32 G memory, we were still only able to fit the SPM (4) with 15000 ($\approx 25\%$) participants. These participants were drawn randomly. The results are given in Figure 11. The age effect in the missing process is clearly non-linear. The other continuous variables, although not perfectly linear, are much closer to being linear.

Therefore we chose to model all variables linearly except for age in the dropout process, which we model as an additive effect.

D Parameter estimates

The parameter estimates from the SPM and naive model introduces in Section 3.2 fitted on the HUNT2 cohort, together with their 95% equitailed credible intervals model are presented in Table 6.

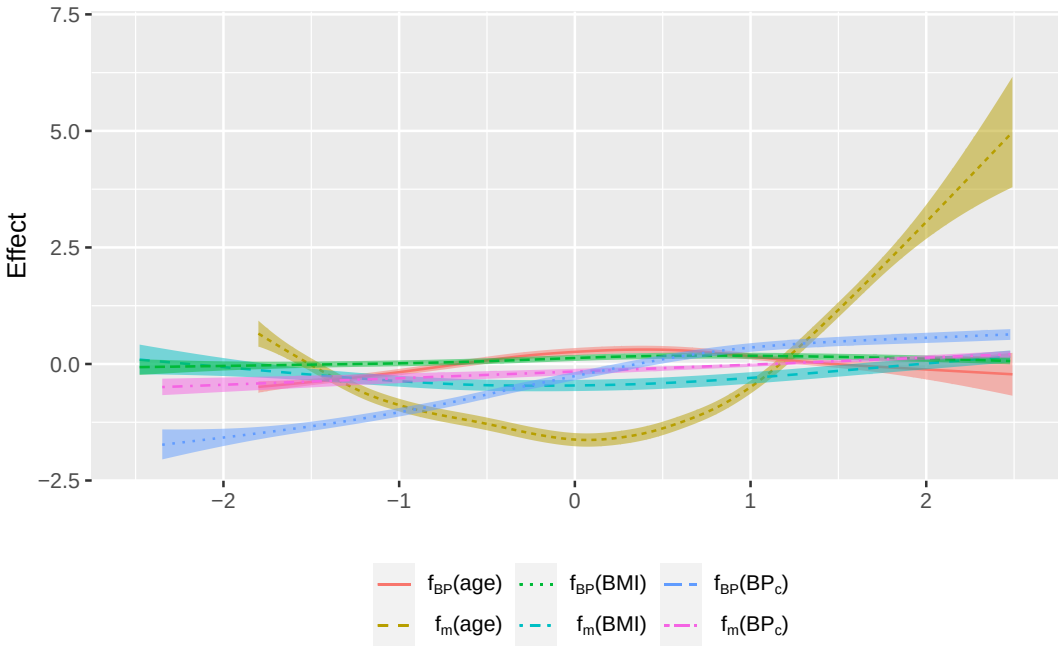


Figure 11: Additive effects of age, BMI, and BP_I in the SPM for both the BP process and the missing process fitted on 15000 randomly drawn participants from the HUNT2 cohort.

Table 6: Summary of all parameters estimated for both the SPM and the naive model fitted on teh HUNT2 cohort. The posterior means and 95% credible intervals are displayed.

	SPM		**Naive**	
	Posterior mean	CI	Posterior mean	CI
α_0	0.275	(0.259, 0.291)	0.134	(0.123, 0.145)
α_{age}	0.244	(0.232, 0.255)	0.246	(0.235, 0.258)
α_{BMI}	0.073	(0.064, 0.082)	0.071	(0.062, 0.080)
α_{BP}	0.599	(0.589, 0.610)	0.578	(0.567, 0.588)
α_{sex}	0.041	(0.025, 0.057)	-0.022	(-0.064, 0.019)
β_0	0.567	(0.475, 0.666)	0.55	(0.460, 0.651)
β_{BMI}	0.087	(0.087, 0.106)	0.081	(0.063, 0.099)
β_{BP}	0.139	(0.139, 0.162)	0.133	(0.111, 0.155)
β_{sex}	0.292	(0.292, 0.329)	0.275	(0.240, 0.310)
σ_{age}	1.412	(0.946, 2.130)	1.400	(0.927, 2.123)
σ_ϵ	0.790	(0.783, 0.797)	0.77	(0.765, 0.776)
c	0.705	(0.645, 0.765)	-	-

E Prior Sensitivity Analysis

Model name	c_0_1	c_0_10	c_0_100	c_1_1	c_1_10	c_1_100	c_10_100
Prior for c	$N(0,1^2)$	$N(0,10^2)$	$N(0,100^2)$	$N(1,1^2)$	$N(1,10^2)$	$N(1,100^2)$	$N(10,100^2)$

Table 7: Priors for association parameter c.

As the SPM model specified in Section 3.2 is rather complex, we perform a sensitivity analysis to evaluate if the model is sensitive to the choice of prior for the association parameter c (3). This parameter is of special interest as it defines the connection between the dropout process and the measurements process, BP_F. We fit the model defined in (3) with different choices for the prior for c displayed in Table 7.

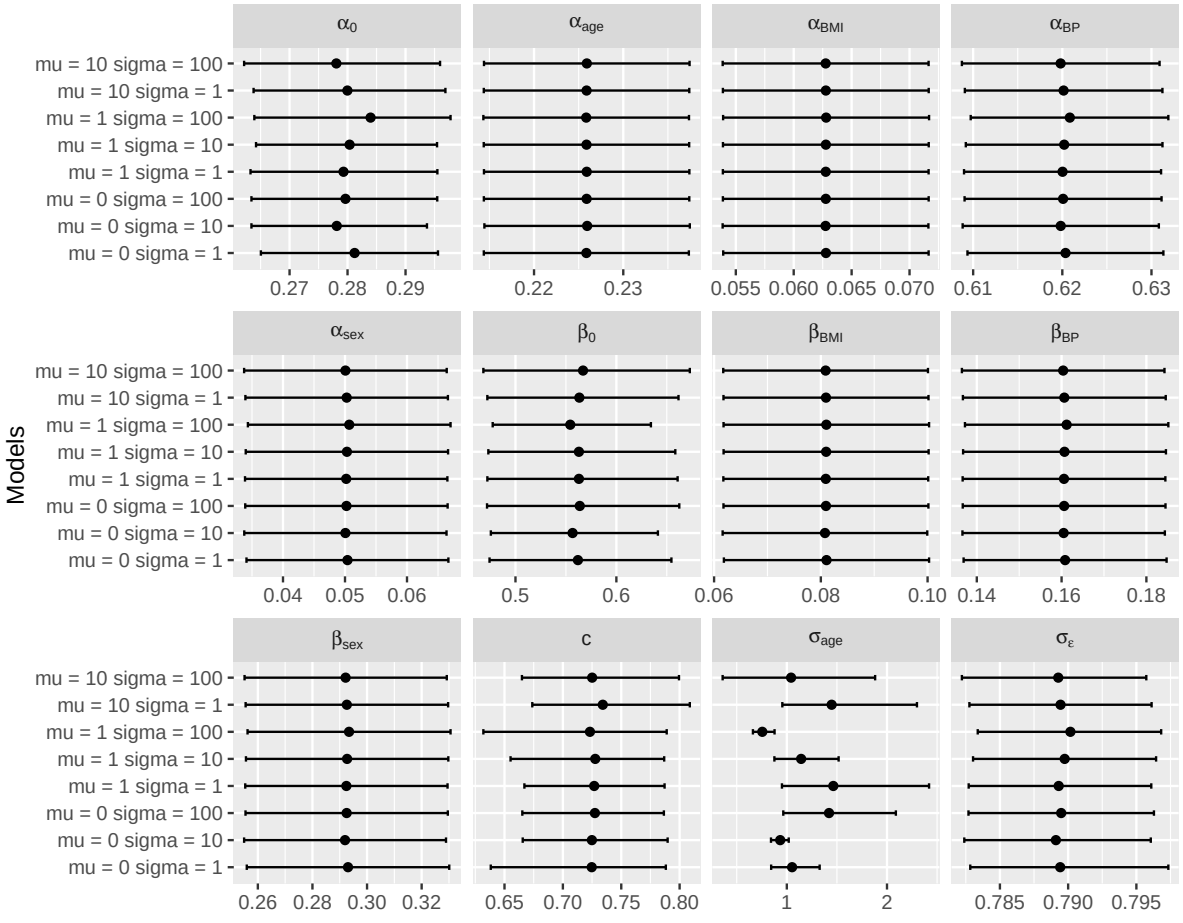


Figure 12: Posterior mean and 95 % credible intervals for all parameters of the SPM with different priors for the association parameter c. The model names are constructed so the first digit is the mean and the second the is the standard deviation used in the prior for c.

Figure 12 shows the resulting 95% equi-tailed credible intervals for all parameters. We clearly see that the latent field, α_0, α_{BP}, α_{age}, α_{BMI}, α_{sex}, β_0, β_{BP}, β_{BMI}, β_{sex}, have almost identical credible intervals. The credible intervals for the hyperparameter σ_{age} vary sligtly

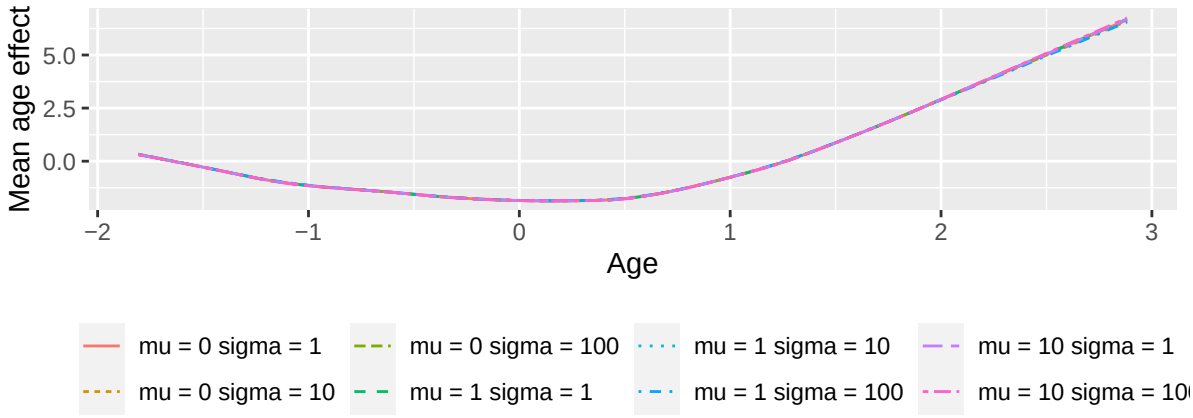


Figure 13: Mean posterior age effect for all models tested in the prior sensitivity study.

more. However, we see in Figure 13 that the resulting posterior mean of the age effect is practically identical for all the different priors. Hence we conclude that the model is not sensitive to different choices of c.

F Supplementary Material for The Simulation Studies on Bias and Coverage

We have summarized the results from the simulation study performed in Section 5.1 exploring the bias and coverage of the SPM and the naive model. In addition we have performed a similar study with data MAR. The distribution of the posterior means from the simulation study when the true parameters are the posterior mean estimates of the SPM and the data is MAR can be seen in Figure 7. Table 8, and Table 9 display the mean posterior mean, bias, and coverage of the parameter estimates for both the SPM and the naive model for simulated data MNAR and MAR. Further we display the difference in bias for the SPM and naive model $Bias_{SPM} - Bias_{naive}$.

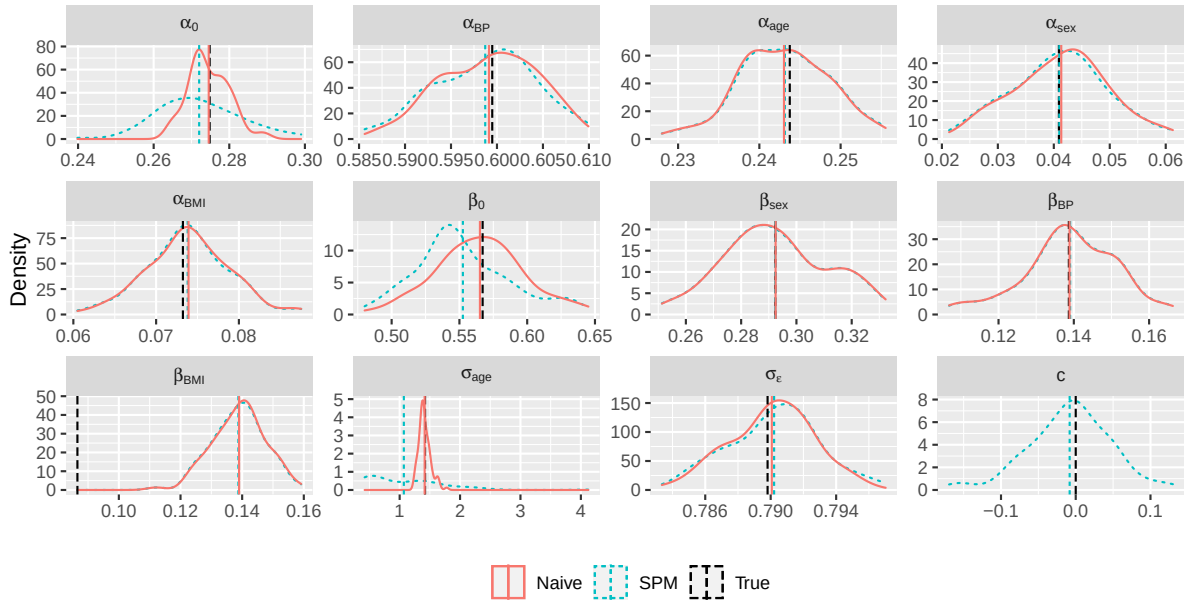


Figure 14: Distribution of posterior mean estimates on simulated data MAR following the SPM. The mean of posterior means for both the SPM and naive model and the true value are indicated by the vertical lines.

Table 8: A summary of the results where we explore the bias and coverage of the SPM and the naive mode when the true model parameters are known. We display the mean posterior mean, bias of posterior means, and coverage for all parameters. In addition, we display the difference in bias for the SPM and the naive model ($Bias_{SPM} - Bias_{naive}$). The true parameters are the posterior means from the SPM as given in Table 6 and the data is MNAR.

		SPM			**Naive model**			
	True value	Mean	Bias	Coverage	Mean	Bias	Coverage	Difference in bias
α_0	0.27	0.24	-3e-02	0.07	0.13	-0.144	0.00	0.112
α_{BP}	0.60	0.60	-4e-03	0.91	0.58	-0.021	0.01	0.017
α_{age}	0.24	0.24	-8e-04	0.98	0.25	0.002	0.97	0.001
α_{bmi}	0.07	0.07	-1e-03	0.95	0.07	-0.006	0.76	0.004
α_{sex}	0.04	0.04	-3e-03	0.97	0.02	-0.018	0.42	0.015
β_0	0.57	0.56	-1e-02	0.98	0.55	-0.016	0.99	0.004
β_{BMI}	0.09	0.13	5e-02	0.01	0.13	0.042	0.01	-0.005
β_{sex}	0.29	0.29	-5e-03	0.93	0.28	-0.016	0.86	0.011
β_{BP}	0.14	0.14	-3e-03	0.90	0.13	-0.007	0.91	0.004
σ_{age}	1.42	1.22	-2e-01	0.67	1.34	-0.082	1.00	-0.119
σ_ϵ	0.79	0.78	-8e-03	0.32	0.77	-0.020	0.00	0.012
c	0.70	0.55	-2e-01	0.00	NA	NA	NA	NA

Table 9: A summary of the results where we explore the bias and coverage of the SPM and the naive model when the true model parameters are known. We display the mean posterior mean, bias of posterior means, and coverage for all parameters. Further, we display the difference in bias for the SPM and the naive model ($Bias_{SPM} - Bias_{naive}$). The true parameters are the posterior means from the SPM as given in Table 6 and the data is MAR meaning $c = 0$.

		SPM			**Naive model**			
	True value	Mean	Bias	Coverage	Mean	Bias	Coverage	Difference in bias
α_0	0.27	0.272	-3e-03	0.8	0.27	-2e-04	1.0	-3e-03
α_{BP}	0.60	0.599	-8e-04	1.0	0.60	-4e-04	1.0	-4e-04
α_{age}	0.24	0.243	-6e-04	1.0	0.24	-7e-04	1.0	1e-04
α_{bmi}	0.07	0.074	6e-04	0.9	0.07	7e-04	0.9	1e-04
α_{sex}	0.04	0.041	7e-05	0.9	0.04	4e-04	1.0	4e-04
β_0	0.57	0.553	-1e-02	1.0	0.57	-2e-03	1.0	-1e-02
β_{BMI}	0.09	0.139	5e-02	0.0	0.14	5e-02	0.0	2e-04
β_{sex}	0.29	0.292	-4e-05	0.9	0.29	9e-05	0.9	4e-05
β_{BP}	0.14	0.139	3e-04	0.9	0.14	2e-04	0.9	-1e-04
σ_{age}	1.42	1.069	-4e-01	0.4	1.41	-6e-03	1.0	-3e-01
σ_ϵ	0.79	0.790	4e-04	1.0	0.79	3e-04	1.0	-1e-04
c	0.00	-0.008	-8e-03	0.9	NA	NA	NA	NA

References

Paul S Albert and Dean A Follmann. Modeling repeated count data subject to informative dropout. *Biometrics*, 56(3):667–677, 2000.

Gunnar H Anderson Jr, Nancy Blakeman, and DH Streeten. The effect of age on prevalence of secondary forms of hypertension in 4429 consecutively referred patients. *Journal of hypertension*, 12(5):609–615, 1994.

Bjørn Olav Åsvold, Arnulf Langhammer, Tommy Aune Rehn, Grete Kjelvik, Trond Viggo Grøntvedt, Elin Pettersen Sørgjerd, Jørn Søberg Fenstad, Oddgeir Holmen, Maria C Stuifbergen, Sigrid Anna Aalberg Vikjord, et al. Cohort profile update: The hunt study, norway. *medRxiv*, 2021.

Narayanaswamy Balakrishnan. *Methods and applications of statistics in the life and health sciences*. John Wiley & Sons, 2009.

Clarice D Brown, Millicent Higgins, Karen A Donato, Frederick C Rohde, Robert Garrison, Eva Obarzanek, Nancy D Ernst, and Michael Horan. Body mass index and the prevalence of hypertension and dyslipidemia. *Obesity research*, 8(9):605–619, 2000.

An Creemers, Niel Hens, Marc Aerts, Geert Molenberghs, Geert Verbeke, and Michael G Kenward. A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal*, 52(1):111–125, 2010.

Peter Diggle and Michael G Kenward. Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43(1):49–73, 1994.

Craig K Enders. Missing not at random models for latent growth curve analyses. *Psychological methods*, 16(1):1, 2011.

Lars Fredrik Espeland. A shared parameter model accounting for non-ignorable missing data due to dropout: Modelling of blood pressure based on the hunt study. Master's thesis, Norwegian University of Science and Technology, 7 2020.

Ludwig Fahrmeir, Thomas Kneib, Stefan Lang, and Brian Marx. *Regression*. Springer, 2007.

Dean Follmann and Margaret Wu. An approximate generalized linear model with random effects for informative missing data. *Biometrics*, 51(1):151–168, 1995.

Ahmed M Gad and Nesma MM Darwish. A shared parameter model for longitudinal data with missing values. *American journal of applied Mathematics and Statistics*, 1(2):30–35, 2013.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

Virgilio Gómez-Rubio. *Bayesian inference with INLA*. Chapman & Hall/CRC Press. Boca Raton, FL., 2020.

James J Heckman. Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 47(1):153–161, 1979.

Robin Henderson, Peter Diggle, and Angela Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.

Aurora Christine Hofman. A-spm-accounting-for-data-mnar, 2021. URL `https://github.com/AuroraSmil/A-SPM-accounting-for-data-MNAR`.

Chanelle J Howe, Stephen R Cole, Bryan Lau, Sonia Napravnik, and Joseph J Eron Jr. Selection bias due to loss to follow up in cohort studies. *Epidemiology (Cambridge, Mass.)*, 27(1):91, 2016.

Shu-Zhong Jiang, Wen Lu, Xue-Feng Zong, Hong-Yun Ruan, and Yi Liu. Obesity and hypertension. *Experimental and therapeutic medicine*, 12(4):2395–2399, 2016.

Niko A Kaciroti and Roderick JA Little. Bayesian sensitivity analyses for longitudinal data with dropouts that are potentially missing not at random: A high dimensional pattern-mixture mode. *Statistics in Medicine*, 40(21):4609–4628, 2021.

S Krokstad, A Langhammer, K Hveem, TL Holmen, K Midthjell, TR Stene, G Bratberg, J Heggland, and J Holmen. Cohort profile: the hunt study, norway. *International journal of epidemiology*, 42(4):968–977, 2013.

Lewington et al. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *The Lancet*, 360(9349):1903–1913, 2002.

Antonio R Linero and Michael J Daniels. Bayesian approaches for missing not at random outcome data: The role of identifying restrictions. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 33(2):198, 2018.

Roderick JA Little. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125–134, 1993.

Roderick JA Little. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the american statistical association*, 90(431):1112–1121, 1995.

Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.

Sara Martino and Andrea Riebler. Integrated nested laplace approximations (inla). *Wiley StatsRef: Statistics Reference Online*, pages 1–19, 2019.

Karthika Mohan and Judea Pearl. Graphical models for processing missing data. *Journal of the American Statistical Association*, 116(534):1023–1037, 2021.

Geert Molenberghs, Caroline Beunckens, Cristina Sotto, and Michael G Kenward. Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2): 371–388, 2008.

Christopher JL Murray, Aleksandr Y Aravkin, Peng Zheng, Cristiana Abbafati, Kaja M Abbas, Mohsen Abbasi-Kangevari, Foad Abd-Allah, Ahmed Abdelalim, Mohammad Abdollahi, Ibrahim Abdollahpour, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet*, 396(10258):1223–1249, 2020.

Erik P Pulkstenis, Thomas R Ten Have, and J Richard Landis. Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. *Journal of the American Statistical Association*, 93(442):438–450, 1998.

R-INLA. R-inla project, 2021. URL `https://www.r-inla.org/home`.

Eleni Rapsomaniki, Adam Timmis, Julie George, Mar Pujades-Rodriguez, Anoop D Shah, Spiros Denaxas, Ian R White, Mark J Caulfield, John E Deanfield, Liam Smeeth, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1· 25 million people. *The Lancet*, 383 (9932):1899–1911, 2014.

Christopher H Rhoads. Problems with tests of the missingness mechanism in quantitative policy studies. *Statistics, Politics, and Policy*, 3(1), 2012.

Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. CRC press, 2005.

Håvard Rue, Sara Martino, and Nicolas Chopin. 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, 2009.

Håvard Rue, Andrea Riebler, Sigrunn H Sørbye, Janine B Illian, Daniel P Simpson, and Finn K Lindgren. Bayesian computing with inla: a review. *Annual Review of Statistics and Its Application*, 4:395–421, 2017.

Ingelin Steinsland, Camilla Thorrud Larsen, Alexandre Roulin, and Henrik Jensen. Quantitative genetic modeling and inference in the presence of nonignorable missing data. *Evolution*, 68(6):1735–1747, 2014.

R Thomas, Ten Have, Allen R Kunselman, Erik P Pulkstenis, and J Richard Landis. Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics*, 54(1):367–383, 1998.

Martin D Tobin, Nuala A Sheehan, Katrina J Scurrah, and Paul R Burton. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Statistics in medicine*, 24(19):2911–2935, 2005.

Masahiko Tozawa, Kunitoshi Iseki, Chiho Iseki, Kozen Kinjo, Yoshiharu Ikemiya, and Shuichi Takishita. Blood pressure predicts risk of developing end-stage renal disease in men and women. *Hypertension*, 41(6):1341–1345, 2003.

Edward F Vonesh, Tom Greene, and Mark D Schluchter. Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in medicine*, 25(1):143–163, 2006.

Paul K Whelton. Epidemiology of hypertension. *Lancet (London, England)*, 344(8915): 101–106, 1994.

World Health Organizatoin. Global action plan for the prevention and control of ncds 2013-2020, 2013. URL `https://www.who.int/publications/i/item/9789241506236`.

Margaret C Wu and Raymond J Carroll. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44(1):175–188, 1988.

Bin Zhou, James Bentham, Mariachiara Di Cesare, Honor Bixby, Goodarz Danaei, Melanie J Cowan, Christopher J Paciorek, Gitanjali Singh, Kaveh Hajifathalian, James E Bennett, et al. Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19· 1 million participants. *The Lancet*, 389(10064):37–55, 2017.

Aurora Christine Hofman

A shared Parameter Model Accounting for Dropout Not at Random

NTNU
Norwegian University of
Science and Technology