UiT The Arctic University of Norway

Department of Mathematics and Statistics

Geographical study of the stroke incidence and mortality rates using Bayesian analysis

Li-Wei Janice Shu

Master's thesis in Statistics STA-3900

September 2023

Abstract

There is large geographical variation in the stroke incidence rate worldwide. In addition, stroke is one of the diseases with highest total mortality rate globally. In this thesis, we examine if there are any geographical variation in stroke incidence in Norway. We first perform a preliminary analysis using classical hypothesis tests within the frequentist framework. Then we enhance the study by using a spatial model. This is done through Integrated Nested Laplace Approximation (INLA) within the Bayesian framework. This method allows us to conveniently incorporate the hierarchical nature of the data structure and look for spatial effects on the stroke incidence rate. Based on the spatial model, we see that aside from two hospital referral areas (OUS and UNN), there are no geographical variation in the stroke incidence in Norway. We further look at stroke mortality by using Bayesian survival analysis. For the study of mortality, we examine for geographic variation as well as effects of socioeconomic status. We run the Cox proportional hazards model and include factors such as geographic, demographic and socioeconomic information. We find that there are no geographical variation in the stroke mortality rate in Norway. However, those with higher income and education level have lower mortality rate.

Acknowledgement

It has been a fun and fulfilling journey in pursuing a master's degree. I have acquired much knowledge, made new friends, and even learned something about myself along the way. This thesis was not done alone, and there are many people behind this that made it possible. I would like to take this opportunity to express my heartfelt appreciation.

First and foremost, I would like to thank my supervisor Sigrunn Holbek Sørbye for her tireless support and guidance. Thank you for always being available and for your patience and invaluable input during the writing process. I have enjoyed working with you immensely and am truly grateful for having you as a mentor.

I would like to thank SKDE for encouraging me to embark on this endeavor and providing me with the opportunity to enrich my competency. Moreover, the complete and rich data source provided by SKDE has enabled the interesting analysis and results.

To my dearest husband Dennis and children Adrian and Justin, thank you for being the strong pillars in holding the family together and steady while I undertook this extra role as a student. You keep me grounded and give me the strength to persevere.

Last but not least, I would like to dedicate this thesis to my mom, Cindy. Thank you for your love and support and for always believing in me. You are the cornerstone of everything I do and the inspiration throughout my life.

Contents

1 Introduction

1.1 Background

Stroke, or cerebrovascular accident (CVA), is when there is a loss of blood flow to part of the brain. When the brain cells do not receive the oxygen and the nutrients from the blood, the affected brain tissue start to die. This can happen in a matter of minutes and can cause serious brain damage, disability or death. Therefore it is an acute condition and requires medical attention immediately (Centers for Disease Control and Prevention (CDC)). There are two types of stroke: ischemic and hemorrhagic. Ischemic stroke is when there is a blockage in the blood vessel hence hindering the blood to flow to the rest of the brain. This is the most common type of stroke and accounts for about 85% of the stroke incidence (The Organization for Economic Cooperation and Development (OECD)). Hemorrhagic stroke is when there is a rupture in a blood vessel and causes bleeding into the brain. While this is less common, it carries very high mortality rate at about $51 - 65\%$ depending on the location of the hemorrhage (Rymer, 2011).

Stroke is the second-leading cause of death and disability worldwide. In 2013, it accounted for over 10% of total death globally (Benjamin et al., 2017). Not only does it have high mortality rate, but the high morbidity results in more than 50% of survivors with life-long disability. Therefore, it puts a heavy burden on the healthcare system and causes great economic and social consequences (Donkor, 2018). There are many risk factors for stroke, such as high blood pressure, high cholesterol, heart disease and diabetes, just to name a few. Some of these may be hereditary, but many are modifiable and are affected by lifestyle and environment. Therefore it is not unthinkable that there are geographical variation for stroke. Globally, there is a more than ten-fold difference between the countries with the highest and lowest age-standardized stroke incidence and mortality rate (Kim et al., 2015). In addition, socioeconomic status has also been shown to contribute to the variation in stroke incidence and mortality (Addo et al., 2012).

In Norway, there were about 8 500 stroke cases in 2014 reported into the Norwegian stroke register. 46% of the patients are women with the mean age of 77.7 and 54% are men with the mean age of 71.9 (Indredavik et al., 2015). The healthcare atlas for the elderly in Norway showed that for the population over the age of 75, there are little geographical variation in the number of inpatients during 2013 - 2015 (Center for Clinical Documentation and Evaluation (SKDE), 2017). This could be an indication that the incidence rate is also evenly spread throughout the country.

Understanding the geographic and demographic patterns of the stroke incidence and mortality helps the healthcare system with planning and providing necessary care to the patients. Furthermore, educating the public with stroke prevention and management can lighten the load that the disease puts on the society.

1.2 Data source

The data for the analysis is based on information from various Norwegian health registries. Stroke patients are identified from the Norwegian stroke register (NHR) during the years 2014-2018. According to the annual report from NHR (Indredavik et al., 2015; Ellekjær et al., 2016; Fjærtoft et al., 2017, 2018, 2019), only 84-87% of the stroke patients are captured in the registry during the study period. In order to ensure that we include as many patients as possible, and more importantly to minimize any geographical skew, the data is supplemented by identifying additional patients with stroke diagnosis from the Norwegian patient register (NPR). Stroke diagnosis is defined as patients with main diagnosis codes as one of the following: I61, I63, I64 [1]

In addition to incidence, we also want to examine the mortality. NHR includes the patients' condition when they were in contact with the specialist health services, such as the type of stroke (hemorrhagic or ischemic), and whether they had previous stroke. Furthermore, the demographics of these patients from Statistics Norway (SSB) is also available for us to identify their socio-economic status such as level of income and education. Lastly, these patients are matched with the Norwegian cause of death registry (DÅR) so we know the date and cause of death for those that died between 2014-2020. Note that the study of the mortality will be based on only patients from NHR and not the additional patients from NPR since the data from SSB and DÅR is available for only the patients from NHR. However, since this still covers the majority of the population and due to the richness of the data, we believe that it is sufficient for the purpose of this thesis.

Data manipulation and visualization will be done in SAS, while data analysis and additional visualization / presentation will be done in R[2] and R-INLA [3].

[1] ICD-10 code I61 is for Nontraumatic intracerebral hemorrhage, I63 is for cerebral infarction, and I64 is for stroke, not specified as hemorrhage or infarction.

[2] https://www.r-project.org/

[3] https://www.r-inla.org/

1.3 Objectives and outline

The objectives of this thesis are to study if there are any spatial pattern for stroke incidence rates, as well as if there are any demographical and geographical differences in stroke mortality rate.

While the subject of stroke incidence and mortality rate is well-studied, there has been few that use Bayesian methodology. In chapter 2, we introduce the fundamentals of the Bayesian method. Furthermore, we present a class of statistical models called latent Gaussian models. This is a very wide class of models that fits into the computational framework known as the integrated nested Laplace approximation (INLA) which simplifies the implementation of Bayesian models.

Chapter 3 focuses on the spatial analysis of the stroke incidence rate. Section 3.1 presents the data from stroke patients based on their place of residence in Norway. Section 3.2 outlines hypothesis tests that will be used under the frequentist framework while section 3.3 discusses the spatial models that will be used within INLA to study the geographical differences. Section 3.4 presents the results of the analysis and lastly section 3.5 provides a summary of the findings.

The survival analysis is presented in chapter 4 in a very similar manner as the spatial analysis above. Section 4.1 presents the mortality data of stroke patients. Section 4.2 looks at the theory behind survival analysis while section 4.3 shows how it is implemented in INLA. Section 4.4 then presents the results of the analysis and 4.5 summarizes the models.

Lastly, a short discussion is included in chapter 5 as concluding remarks for the thesis.

2 Bayesian statistics

In this chapter we will introduce the theory behind Bayesian inference by first examining Bayes' theorem and how it is applied to Bayesian inference. We will then look at the components of the inference, namely the prior distribution, likelihood function and posterior distribution.

The second part of this chapter is on integrated nested Laplace approximation. It is a computational technique that can be used when working with Bayesian inference. We will look at the specifications and assumptions with this approach and how it is done.

2.1 Bayesian inference

Bayesian inference is a branch of statistical analysis tool that uses probability theory to obtain a distribution about the parameters by updating prior knowledge with observed data. In the Bayesian framework, the parameters are unknown and treated as random variables. They can be expressed as probability distributions which are used for the inference.

One of the advantage with this method is that it can easily be adapted to incorporate hierarchical models. This allows us to analyse the information that may occur at different levels and capture the dependencies and variations between them. It is a powerful tool that is very versatile in dealing with complex data.

2.1.1 Bayes' theorem

The cornerstone of Bayesian inference is Bayes' theorem (Bayes, 1763). Suppose A and B are events on a sample space, then

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}. \tag{2.1}$$

This can be used for the probability distributions of the unknown parameters. For the observed data $y^T = (y_1, \ldots, y_n)$ and the m unknown parameters $\theta^T = (\theta_1, \ldots, \theta_m)$, the joint probability distribution can be expressed as the product of the probability distribution of y given θ and the probability distribution of θ. So,

$$\pi(\theta, y) = \pi(y|\theta)\,\pi(\theta). \tag{2.2}$$

When Bayes' theorem in equation (2.1) is applied to the data and parameters in equation (2.2), we end up with a conditional probability distribution of the parameters given the data

$$\pi(\theta|y) = \frac{\pi(\theta, y)}{\pi(y)} = \frac{\pi(y|\theta)\,\pi(\theta)}{\pi(y)}.$$

Since y represent the data, $\pi(y)$ does not depend on θ and can be viewed as a normalizing constant. We therefore omit the denominator and consider the simpler form

$$\pi(\theta|y) \propto \pi(y|\theta)\,\pi(\theta). \tag{2.3}$$

$\pi(\theta|y)$ on the left side of expression (2.3) is ultimately what we are after in Bayesian inference: probability distribution of the parameters given data, which is also referred to as the posterior distribution. If we examine how the posterior distribution can be derived, we have $\pi(y|\theta)$ which can be expressed as a function of θ instead of y. Then we have $l(\theta|y)$, which is the likelihood function of the parameters given data. The other component is $\pi(\theta)$ which represents the probability distribution of the parameters. This captures our initial beliefs of the parameters and is called the prior distribution.

In all, we can say that Bayesian inferences can be made with the knowledge of the likelihood function and the prior distribution. We update our previous beliefs about the parameter with the observed data through the likelihood function. This allows us to arrive at the posterior distribution for the parameters which can be used to make prediction, estimation, or draw conclusion about the data.

2.1.2 Prior distribution

The priors are the distributions of the unknown parameters θ without the knowledge of data. As it is a component in determining the posterior distribution, the choice of prior is important and affects the outcome of the model. There are many different types of prior, for example, non-informative, weakly informative, or informative priors.

If we don't know much about the parameter, we can use a non-informative, vague or flat prior. This means that we assume very little about the parameter and allow the data to be the main driver in determining the posterior. Some examples of non-informative priors are: flat prior where we assign the same probability or density to all values in the possible range, Gaussian prior with large variance to allow for uncertainty, or Jeffrey's prior which is invariant to parameter transformation (Jeffreys, 1946; Blangiardo and Cameletti, 2015).

On the other end of the spectrum, we have also informative priors. It is when the distribution of θ is known, or that we have reasonable beliefs based on, for instance, previous knowledge, expert opinion or the population from which the parameters were drawn. Having an informative prior ensures that the model has as much information as possible, and is especially useful if we have limited data. There is a subclass of priors which are called conjugate and non-conjugate priors. Conjugate priors are such that the resulting

posterior falls under the same class of distribution as the prior. Consider a prior that has a beta distribution with a binomial likelihood, which results in a posterior that is also of beta distribution. Conjugate priors provide a simpler computational framework, but are not always appropriate or possible (Gelman et al., 2013).

In reality, non-informative priors force us to model with ignorance, and informative priors are not easily obtainable or sometimes too strong that they overpower the data. Weakly informative priors can be a compromise to these drawbacks. It can be built by either starting with a non-informative prior and adding some information or constraints until we arrive at a reasonable posterior. Alternatively, one can also start with an informative prior and relax the beliefs to allow for some uncertainty (Gelman et al., 2013). Choosing a prior is often not easy or clear-cut and may require some experimentation and evaluation of the results. Later on in section 2.2.4 we will look at a specific type of weakly informative prior called penalized complexity prior.

2.1.3 Likelihood function

The likelihood function is the probability distribution of observing data y given the parameter θ. It is a function of θ and can be expressed as

$$l(\theta|y) = \pi(y|\theta).$$

It captures how well the parameter explains the data. In Bayesian inference, the observed data is expressed through the likelihood function which is in turn used to update the prior distribution in order to obtain the posterior distribution.

2.1.4 Posterior distribution

If we have several parameters in the model, then the posterior is called a joint posterior distribution. When interpreting the model, we may want to be concerned with only a subset of them at a time. The marginal posterior distribution of the parameters of interest can be obtained by integrating the joint posterior over the parameters that are not of interest, so-called nuisance parameters (Gelman et al., 2013). This can further be written as a product of the conditional posterior distribution given the nuisance parameter and the marginal posterior distribution of the nuisance parameter. To see this mathematically, consider $\theta^T = (\theta_{(1)}, \theta_{(2)})$ where $\theta_{(1)}$ are the parameters of interest and $\theta_{(2)}$ are the nuisance parameters, then we have

$$\begin{aligned}\pi(\theta_{(1)}|y) &= \int \pi(\theta_{(1)}, \theta_{(2)}|y)\, d\theta_{(2)} \\ &= \int \pi(\theta_{(1)}|\theta_{(2)}, y)\, \pi(\theta_{(2)}|y)\, d\theta_{(2)}. \end{aligned} \tag{2.4}$$

Inferences from the marginal posterior distributions can then be summarized by point estimates and credible intervals. Posterior distributions are probability distributions, and the summary statistics is therefore quite

straight forward. The most commonly used point estimator is the posterior mean. It can be written as

$$E(\theta|y) = \int \theta\pi(\theta|y)d\theta.$$

Since we are working with a distribution, we can also easily find the median and the quantiles. The posterior median $\theta_{0.5}$ is the value that divides the distribution into two equal parts (Blangiardo and Cameletti, 2015):

$$P(\theta \leq \theta_{0.5}) = 0.5 \text{ and } P(\theta \geq \theta_{0.5}) = 0.5.$$

Similarly we can define the 2.5% and 97.5% quantiles as

$$P(\theta \leq \theta_{0.025}) = 0.025 \text{ and } P(\theta \geq \theta_{0.975}) = 0.025,$$

which gives us the end values of the 95% equi-tail credible interval. Since the summary statistics are obtained from the probability distribution, the interpretation is intuitive. A 95% credible interval simply means that there is 95% probability that the value of θ is within the interval. In frequentist statistics we use the confidence interval and the alpha level to interpret significance. In Bayesian statistics, there is not such notion of significance, but we often use credible interval as a guideline and say that if the interval covers 0, then it indicates that there is no positive or negative effect, hence can be thought of as no significance.

2.1.5 Hierarchical models

The parameters in the model can in some cases be related to each other. Bayesian inference allows for this dependency through hierarchical models (Blangiardo and Cameletti, 2015). It incorporates complex data that are structured in different hierarchies into the Bayesian probabilistic framework and effectively accounts for the variability within and between the levels. In addition to the data y and parameters θ as we have seen so far, hierarchical model introduces hyperparameters ϕ to describe the dependencies between the θs. One example would be patients in different hospitals, where the variability between the patients is described by the parameter θ and the variability between the hospitals is captured by the hyperparameter ϕ. The joint posterior distribution in equation (2.3) is extended to reflect the extra level of information and becomes

$$\pi(\theta, \phi|y) \propto \pi(y|\theta, \phi)\, \pi(\theta|\phi)\, \pi(\phi). \tag{2.5}$$

This means that hierarchical models are a three-stage process: first identify the likelihood of the parameters $l(\theta|y)$, then the prior of the parameters given the hyperparameters $\pi(\theta|\phi)$, and lastly prior of the hyperparameters $\pi(\phi)$. Once we have the joint posterior distribution $\pi(\theta, \phi|y)$, similar to the process described in equation (2.4), we want to find the marginal posterior of the parameters of interest, θ, by integrating out the nuisance hyperparameters, ϕ.

As we see, hierarchical models can easily be adapted to the Bayesian framework which gives us a very powerful tool in statistical analysis. Through the additional level of the hyperparameter that describes the parameter, the model borrows information from similar populations and reduces the uncertainty in each population. This results in a model that is more robust while still maintaining the variability between the different populations.

2.2 Integrated nested Laplace approximation

2.2.1 Introduction to INLA

Bayesian inference is done through the posterior distribution of the parameters for prediction, estimation and/or conclusion. It is especially powerful when we incorporate hierarchical structure in the models. However, due to the complexity of the models, this often means that obtaining the distribution analytically is difficult or sometime impossible. Simulation-based techniques such as Markov chain Monte Carlo (MCMC) are popular choices to explore the target distribution. The property of Markov chain states that the conditional probability of the next state is dependent solely on the current state. One may use a sampling technique such as Gibbs sampling to construct a chain such that at its stationary state, the approximate distribution is equal to the desired target distribution (Gelman et al., 2013).

While sampling makes it possible to estimate the posterior distribution, it can be computationally expensive and time consuming. In the case of hierarchical models, due to the richness of the data and/or the complex model structure, this may still take hours or days even with modern powerful computers. Rue, Martino and Chopin introduced a method called integrated nested Laplace approximations (INLA) in 2009 for approximating the posterior distributions for a particular type of model, namely the latent Gaussian model (Rue et al., 2009). It improves the computational efficiency significantly without sacrificing accuracy (Rue et al., 2017).

Latent Gaussian model

To understand what a latent Gaussian model (LGM) is, we first consider a general form of an additive model

$$\eta_i = g(\mu_i) = \beta_0 + \sum_{j=1}^{n_1} \beta_j z_{ji} + \sum_{k=1}^{n_2} f_k(c_{ki}), \quad i = 1, \ldots, n, \tag{2.6}$$

where $\mu_i = E(y_i)$, $g(.)$ is a link function, β_0 is the intercept, β_j are the parameters for the covariates z_{ji}, and f_k are non-linear functions, smoothing splines, or some random effects (Rue et al., 2017). These terms are

all random variables under the Bayesian framework. Consider the vector x as a latent field from equation (2.6) for which each of the parameters have a Gaussian prior. Then we have

$$x = \{\beta_0, \beta_1, ..., \beta_{n_1}, f_1, ..., f_{n_2}, \eta_1, ..., \eta_n\}, \tag{2.7}$$

and the model is called a latent Gaussian model. The latent components fs have priors with parameters ϕ. A hierarchical model as in equation (2.5) that falls under the subclass of LGM, where the data y is conditionally independent given the latent field x, and hyperparameters ϕ, can be formulated as

$$\begin{aligned} y|x, \phi_1 &\sim \prod_{i \in I} \pi(y_i|x_i, \phi_1) \\ x|\phi_2 &\sim \mathcal{N}(\mu(\phi_2), Q^{-1}(\phi_2)) \\ \phi &\sim \pi(\phi), \end{aligned}$$

where Q is the precision matrix. Note that the dimension of ϕ is generally small, but the dimension of x is often large as it may be necessary to capture the complexity and the variation in the data, especially in a hierarchical model (Rue et al., 2017).

Gaussian Markov random field

Gaussian Markov random field (GMRF), as the name suggests, is a random vector that has a multivariate Gaussian distribution and follows the Markov property. Markov property is such that some components in the vector is conditionally independent, i.e. $x_i \perp x_j|x_{-ij}$, for $i \neq j$ (Rue and Held, 2005). Rue et al. (2009) showed that this property is encoded in the precision (inverse of covariance) matrix Q such that $x_i \perp x_j|x_{-ij} \Leftrightarrow Q_{ij} = 0$. A LGM is a GMRF if many components in the latent field can be assumed to be conditionally independent, which gives us the sparse precision matrix. This is the key attribute that enables the computational efficiency in INLA (Rue et al., 2009, 2017).

Laplace approximation

Laplace approximation can be used to approximate an integral. Assume $f(x)$ is non-negative and integrable. We want to approximate the integral

$$I = \int_x \exp(nf(x))dx.$$

Consider the Taylor expansion of $f(x)$ around x_0 where f(x) is maximize, then we have

$$\begin{aligned} I &\approx \int_x \exp\left[n\left(f(x_0) + (x-x_0)f'(x_0) + \frac{(x-x_0)^2}{2}f''(x_0)\right)\right] dx \\ &= \int_x \exp\left[n\left(f(x_0) + \frac{(x-x_0)^2}{2}f''(x_0)\right)\right] dx \\ &= \exp(nf(x_0)) \int_x \exp\left(n\frac{(x-x_0)^2}{2}f''(x_0)\right) dx \\ &= \exp(nf(x_0)) \int_x \exp\left(-\frac{1}{2}\frac{(x-x_0)^2}{-1/(nf''(x_0))}\right) dx. \end{aligned}$$

Let $\sigma^2 = -1/(nf''(x_0))$, then we have

$$\begin{aligned} I &\approx \exp(nf(x_0))\sqrt{2\pi}\sigma \int_x \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(x-x_0)^2}{\sigma^2}\right) dx \\ &\approx \exp(nf(x_0))\sqrt{2\pi}\sigma \end{aligned}$$

where the integrand is approximated with a Gaussian distribution matching the mode and the curvature at the mode (Rue et al., 2009, 2017).

INLA

With the understanding of LGM, GMRF and Laplace approximation, we can now see how INLA uses these properties to efficiently evaluate the posterior distribution. A LGM is considered a hierarchical model when the f functions have hyperparameters ϕ. The joint posterior follows equation (2.5) and can be expressed as

$$\pi(x, \phi|y) \propto \pi(y|x, \phi)\, \pi(x|\phi)\, \pi(\phi).$$

In INLA, instead of obtaining the joint posterior distribution $\pi(x, \phi|y)$, we are interested in evaluating the marginal posterior distributions for each hyperparameter $\pi(\phi|y)$, and component of the latent field $\pi(x|y)$. The posterior marginals for the hyperparameters are

$$\pi(\phi_j|y) = \int \pi(\phi|y)\, d\phi_{-j} \tag{2.8}$$

and the posterior marginals for the components of the latent field are

$$\pi(x_i|y) = \int \pi(x_i|\phi, y)\, \pi(\phi|y)\, d\phi. \tag{2.9}$$

So what needs to be evaluated are the terms inside the integrals, $\pi(\phi|y)$ and $\pi(x_i|\phi, y)$.

$\pi(\phi|y)$ is needed for the posterior marginals for the hyperparameters, but is also nested in the marginals for the latent field. In order to evaluate $\pi(\phi|y)$, we use probability theory and rewrite it to $\pi(x, \phi, y)/\pi(x|\phi, y)$.

The denominator can be approximated with Laplace approximation by evaluating the Gaussian density at its posterior mode (Rue et al., 2009, 2017). Therefore we have :

$$\tilde{\pi}(\phi|y) \propto \frac{\pi(x, \phi, y)}{\tilde{\pi}_G(x|\phi, y)}\bigg|_{x=x^*(\phi)}, \tag{2.10}$$

where $x^*(\phi)$ is the mode of $\tilde{\pi}_G$ given ϕ.

Similarly, $\pi(x_i|\phi, y)$ can be obtained by the same logic. However, since the dimension of x might be large, it may be more expensive to perform Laplace approximation. Some other methods that can also be considered are Gaussian approximation, Laplace approximation with partitioning of the latent field, or simplified Laplace approximation (Rue et al., 2009, 2017).

Finally, once we have $\tilde{\pi}(\phi|y)$ and $\tilde{\pi}(x_i|\phi, y)$, we are ready to calculate the integrals in equations (2.8) and (2.9). This can be done through numerical integration. For the posterior marginals for the latent field in equation (2.9), we would have

$$\tilde{\pi}(x_i|y) \approx \sum_k \tilde{\pi}(x_i|\phi_k, y)\, \tilde{\pi}(\phi_k|y)\, \Delta_k,$$

where Δ_k denotes the weights. This is possible since the dimension of ϕ is usually small (Rue et al., 2009).

In summary, INLA uses numerical integration on the nested marginals with Laplace approximation to arrive at the desired marginal posterior distributions. It is possible due to the properties of GMRF and the sparse precision matrix which contributes to the computational efficiency. While this method is limited to LGM, it is applicable to a large collection of models such as spatial and spatio-temporal models, time series analysis and survival analysis, to name a few. INLA has gained popularity in the recent years (Van Niekerk et al., 2023) due to the improvement of efficiency over traditional simulation-based techniques while maintaining accuracy.

2.2.2 Independent models for the latent effects

In INLA, all components in the latent field have priors that are Gaussian distributed. The βs in (2.7) have non-informative priors with mean zero and large variance. However, the f functions in equation (2.6), which are also the latent effects in a Bayesian hierarchical model, should have priors that are informative. One of the most common and simplest options is the independent and identically distributed (IID) model. Consider $g(\mu_i) = \beta_0 + u_i$ where u_i denotes the random independent effects. Then we have

$$u_i \sim \mathcal{N}(0, \tau_u^{-1}), \quad i = 1, \ldots, n.$$

where the precision τ_u is the hyperparameter. Since the effects are assumed i.i.d., it can be used to capture unstructured variability in the data.

2.2.3 Random walk for the latent effects

Another common model for the latent effects are the random walk models (Gómez-Rubio, 2020; Blangiardo and Cameletti, 2015). This is often used for covariates that are on a time scale. For the first order random walk (RW1), consider a vector $v = (v_1, \ldots, v_k)$ that is sorted in ascending time order, the first order difference Δv_j can be written as

$$\Delta v_j = v_j - v_{j-1} \overset{\text{iid}}{\sim} \mathcal{N}(0, \tau^{-1}), \quad j = 2, \ldots, k.$$

This therefore gives us a GMRF that is of rank $k-1$, which is improper. We call this an intrinsic GMRF (IGMRF) due to the rank deficiency (Rue and Held, 2005). v_j is independent of $v_1, ...v_{j-2}$ given v_{j-1}. This conditional independence gives us a sparse precision matrix that looks like:

$$Q = \tau R = \tau \begin{pmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{pmatrix}.$$

With a RW1 prior for the latent component, deviation from a constant level is penalized. τ is the hyperparameter which defines the smoothness of the curvature for the random walk model.

The second order random walk (RW2) is constructed in similar ways, but assumes that the second order difference $\Delta^2 v_j$ is Gaussian distributed, so we have

$$\Delta^2 v_j = v_j - 2v_{j-1} + v_{j-2} \overset{\text{iid}}{\sim} \mathcal{N}(0, \tau^{-1}), \quad j = 3, \ldots, k.$$

In a RW2 prior, v_j is independent of the other elements, given v_{j-1} and v_{j-2}. The sparse precision matrix is

$$Q = \tau R = \tau \begin{pmatrix} 1 & -2 & 1 & & & & & & & \\ -2 & 5 & -4 & 1 & & & & & & \\ 1 & -4 & 6 & -4 & 1 & & & & & \\ & 1 & -4 & 6 & -4 & 1 & & & & \\ & & \ddots & \ddots & \ddots & \ddots & & & & \\ & & & 1 & -4 & 6 & -4 & 1 & & \\ & & & & 1 & -4 & 6 & -4 & 1 & \\ & & & & & 1 & -4 & 5 & -2 & \\ & & & & & & 1 & -2 & 1 & \end{pmatrix}.$$

Deviation from a linear trend is penalized therefore giving a smoother result (Rue and Held, 2005).

2.2.4 Penalized complexity prior

The choice of prior is important and can affect the resulting posterior distribution. In equation (2.7) we have a latent field x with Gaussian priors within the LGM framework. However, choosing a prior for the hyperparameters is often not as easy or clear cut. Simpsons et al. proposed a class of priors called penalized complexity (PC) priors (Simpson et al., 2017). PC prior uses a systematic approach by using an inherited nested structure, which contains the base model and a flexible version of the model. It penalizes the complexity induced by deviating from the base model (Simpson et al., 2017; Sørbye and Rue, 2017). Consider hyperparameter ϕ, PC prior is built based on the following four principles:

1. Occam's razor - base model is preferred unless there is enough evidence to suggest otherwise. Let $\pi(x|\phi)$ be the prior distribution, then $\pi(x|\phi = 0)$ is the base model (Simpson et al., 2017).

2. Measure of complexity - use of Kullback-Leibler divergence (KLD) (Kullback and Leibler, 1951) to measure the information lost by using a base model over the more flexible model. KLD is defined as:

$$\text{KLD}(\phi) = \text{KLD}(\pi(x|\phi)||\pi(x|\phi = 0)) = \int \pi(x|\phi) \log\left(\frac{\pi(x|\phi)}{\pi(x|\phi = 0)}\right) dx.$$

   For ease of interpretation, the complexity can be expressed as a unidirectional measure of distance between the two models (Simpson et al., 2017)

$$d(\pi(x|\phi)||\pi(x|\phi = 0)) = \sqrt{2\text{KLD}(\phi)}.$$

3. Constant-rate penalization - the deviation from the base model with distance d is penalized with a constant rate of decay $r \in (0, 1)$ so that the prior satisfies the following

$$\frac{\pi_d(d+\delta)}{\pi_d(d)} = r^\delta, \quad d, \delta \geq 0.$$

This means the change in the prior by adjusting δ is independent of the distance d. Moreover, it implies an exponential density $\pi(d) = \lambda \exp(-\lambda d)$ for $\lambda = -\log(r)$. The PC prior can then be obtained by change of variable transformation (Sørbye and Rue, 2017).

4. User-defined scaling - user defines parameters U and α. For $Q(\phi)$ as a transformation of ϕ, we have $P(Q(\phi) > U) = \alpha$. Here, U controls the upper limit of a tail event and α controls the probability of exceeding the limit (Simpson et al., 2017).

These priors have the advantages of being robust and invariant to reparameterization as well as easy interpretation.

2.2.5 Joint model with two response vectors

The most commonly seen models have one response vector. However, it is possible to set up a joint model with multiple response vectors (Gómez-Rubio, 2020). This allows us to model the responses simultaneously. Consider the response vectors, $\boldsymbol{y} = (y_1, \ldots, y_n)$ and $\boldsymbol{z} = (z_1, \ldots, z_m)$, they can be set up in a matrix in the form of:

$$\begin{pmatrix} y_1 & \text{NA} \\ \vdots & \vdots \\ y_n & \text{NA} \\ \text{NA} & z_1 \\ \vdots & \vdots \\ \text{NA} & z_m \end{pmatrix} \tag{2.11}$$

The covariates can be set up in the same way if they correspond to only one of the response vectors. Consider that we want $\boldsymbol{y}$ and $\boldsymbol{z}$ to have separate intercepts and that covariate $\boldsymbol{u} = (u_1, \ldots, u_n)$ affects $\boldsymbol{y}$,

and $\boldsymbol{v} = (v_1, \ldots, v_m)$ affects $\boldsymbol{z}$. Then the linear predictors would look like:

$$\begin{pmatrix} 1 & \text{NA} & u_1 & \text{NA} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \text{NA} & u_n & \text{NA} \\ \text{NA} & 1 & \text{NA} & v_1 \\ \vdots & \vdots & \vdots & \vdots \\ \text{NA} & 1 & \text{NA} & v_m \end{pmatrix} \tag{2.12}$$

If all covariates affect only one of the response vectors, then the joint model would yield the same result as if the responses were modelled separately. The interesting aspect with a joint model is that there may be shared terms between the different responses. Consider $\boldsymbol{w} = (w_1, \ldots, w_{n+m})$ which affects both $\boldsymbol{y}$ and $\boldsymbol{z}$. It would then be a column in the matrix filled with values, without any NAs like what we see for the other vectors above.

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \\ w_{n+1} \\ \vdots \\ w_{n+m} \end{pmatrix} \tag{2.13}$$

Lastly, note that while the example is for two response vectors, this setup can be generalized to any number of response vectors.

2.2.6 Deviance information criteria

As a part of the model building process, we explore multiple models and choose one in the end as a final model. One common metric used for Bayesian models is the deviance information criteria DIC (Spiegelhalter et al., 2002). DIC is the sum of the measure of fit (deviance) and model complexity. The deviance is defined as

$$\mathrm{D}(\theta) = -2\log(l(y|\theta)).$$

The model complexity is measured with the effective number of parameters and is defined as the posterior mean of the deviance minus the deviance of the posterior mean, and can be written as

$$p_{\mathrm{D}} = \overline{\mathrm{D}(\theta)} - \mathrm{D}(\overline{\theta}).$$

So,

$$\text{DIC} = \overline{\text{D}(\theta)} + p_{\text{D}}.$$

The smaller the deviance, the better the fit, which also usually means higher complexity. We want to minimize both terms, therefore a model with a smaller DIC is preferred.

3 Spatial analysis

In this chapter, we would like to study the stroke incidence rate during the years 2014-2018 for the population of ages 45 and over. The main focus is to explore if there are any geographical variation in the incidence rates. In other words, are there any geographical areas that have different incidence rates than the other areas. If so, is there an underlying pattern that can be explained by a spatial model.

Section 3.1 introduces the data we are working with. Section 3.2 and section 3.3 cover the theory behind the analysis tools: hypothesis tests and spatial models, respectively. Section 3.4 presents the analysis and the results. Lastly in section 3.5, the results from the hypothesis tests and spatial models are compared and a conclusion is drawn.

3.1 Stroke patients and incidence rates

From the Norwegian stroke register (NHR), there are 40 120 stroke patients ages 18-105 during the period 2014 - 2018. In addition, 7 968 stroke patients from NPR in the same time period are also included in the study, which gives us a total of 48 088 patients. Figure 1 (left) shows the distribution of the number of stroke patients by gender and age (in 5-year blocks, hereafter referred to as age blocks or blocks). Below the age of 80, there are more men with stroke than women. However, for the population of 80 years and older, there are more women with stroke comparing to men. This however does not necessarily mean that women are more likely to get stroke than men in the older age groups. This is due to the fact that there are more women in the older population. We therefore look at the average annual percent of population with stroke as shown in figure 1 (right). Now we see that there is a higher proportion of men with stroke comparing to the proportion of women with stroke in the same age block. This remains true throughout the population.

In addition, notice that while there are stroke patients in every age block, there are very few under the age of 45. We will therefore concentrate the analysis on stroke patients between ages 45 and 105, which gives us 45 235 patients, or an annual average of 9 047.

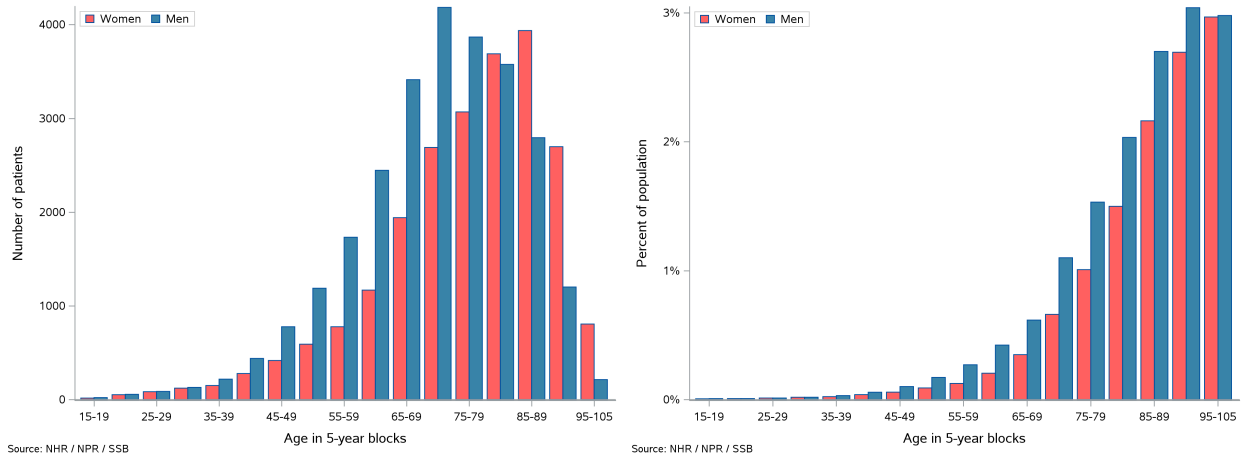

Figure 1: Stroke patients during 2014-2018. Figure on the left shows the average annual number of patients by age and gender while the figure on the right shows the average annual percent of population with stroke by age and gender.

For each of the patients, we have data on the municipality they live in. This information allows us to assign the patients to the areas in which the hospitals serve. There are 21 hospital referral areas (RA) in Norway, and they can be aggregated into 4 regional health authorities (RHA): north, central, west and south-east. To study the geographical variation, we illustrate the stroke incidence rate by the 21 RAs on a Norwegian map in figure 2. The figure on the left shows the average annual crude (unadjusted) incidence rate calculated as number of patients divided by the population in the RA. However since each RA have a different make-up of age and gender, and that these are known risk factors of stroke (Roy-O'Reilly and McCullough, 2018), we apply direct age- and gender-adjustment using the national numbers as the reference. This ensures that the rates across different geographical areas are comparable (Nyi, 2000). Figure 2 (right) shows the age- and gender-adjusted rates across Norway. If we compare the difference between the two maps, we see that some RAs have a lower rate after adjustment. This is due to the fact that the population is older in these areas compared to the national average, so the crude rate appears higher. By the same token, those that have a higher rate after adjustment are the ones with a younger population. This highlights the importance of age- and gender-adjustment in order to be able to compare the incidence rates across areas and regions. We will investigate the data in this section by using the adjusted rate. However, in the analysis in section 3.4, the unadjusted (crude) rate will be used and the difference in the age and gender proportion will be addressed by the appropriate test and model designs.

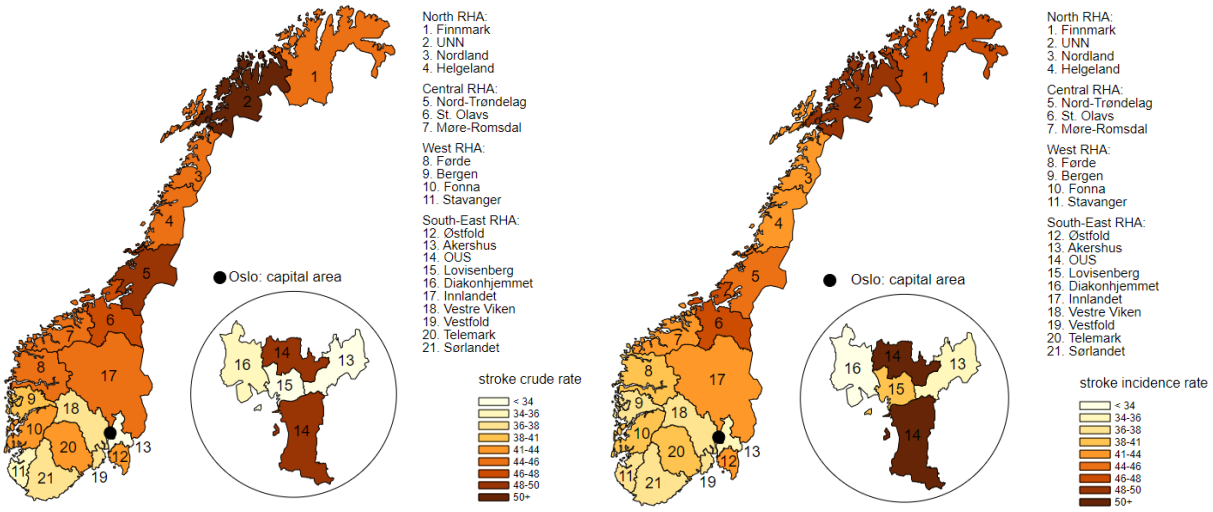

Figure 2: (Left) Unadjusted crude stroke incidence rate. (Right) Age- and gender-adjusted stroke incidence rate. The rates are number of patients per 10 000 population.

Figure 3 (left) presents the same information as the map in figure 2 (right) but with the adjusted incidence rates in descending order. We see that the referral area Oslo university hospital (OUS, area 14) has the highest rate of 52 stroke patients per 10 000 population while referral area Diakonhjemmet (region 16) has the lowest rate of 33 stroke patients per 10 000 population. This gives us a factor of 1.6. In epidemiology, this is often considered little geographical variation. However, it depends on many other factors, such as volume and mix of the population. Figure 3 (right) illustrates the adjusted incidence rate aggregated to the regional level. The north RHA has the highest rate followed closely by the central RHA. Both south-east RHA and west RHA are below the national average and have very similar rates. We will investigate this in section 3.4.1 by doing some hypothesis testing to see if there is statistical significance within the regions (RHA) as well as the referral areas (RA).

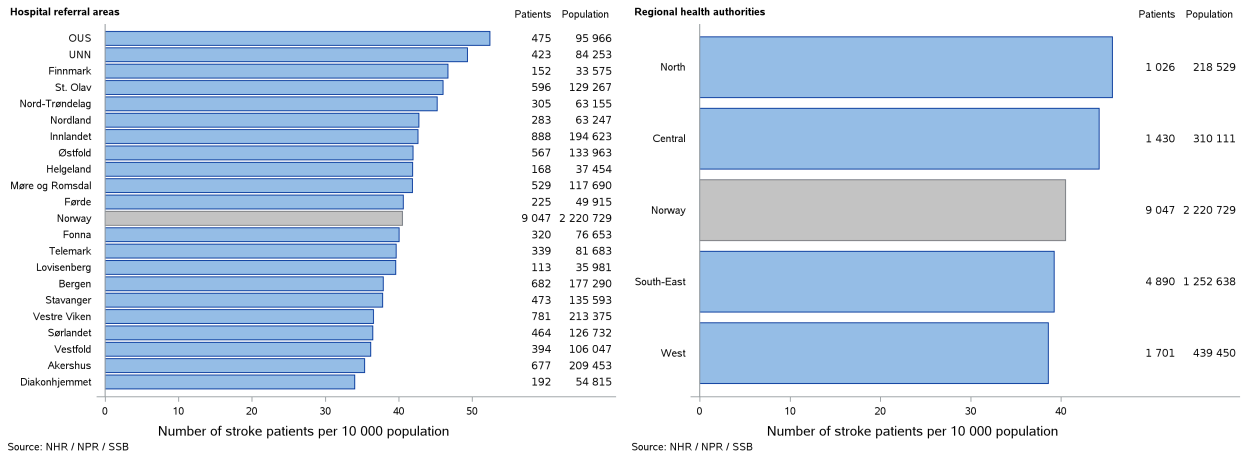

Figure 3: (Left) Age- and gender-adjusted stroke incidence rate per 10 000 population for the 21 hospital referral areas. (Right) The same rates aggregated to the 4 regional health authorities.

3.2 Hypothesis tests

Hypothesis test is used to determine if the differences observed in data is by chance or if there are some factors that may explain the relationship. There are many types of such tests, each suitable for different types of data and scenarios.

3.2.1 Friedman test

The data we are testing on is number of stroke patients in the geographical areas in which we can not assume normality. Thus, we consider non-parametric tests. Moreover, the data is divided into independent age blocks within the the areas (RA/RHA), and there is only one measurement for a given block within an area, the resulting model is referred to as a complete block design. For this we can use the Friedman test for significance between 2 or more groups (Friedman, 1937).

Let y_j be the average number of stroke patients per year in area j and E_j be the population in area j during the reference year, then stroke incidence rate per year per 10 000 population can be estimated by $\hat{\lambda_j} = (y_j/E_j) \times 10\,000$. For $j = 1, \ldots, k$ where k is the number of areas, the hypotheses for the underlying rates using Friedman test are:

$$H_0: \quad \lambda_1 = \lambda_2 = \ldots = \lambda_k$$
$$H_1: \text{At least one of the } \lambda\text{s differ}$$

First we set up the data in a matrix $X = (x_{ij})$ with $i = 1, \ldots, n$ for the blocks and $j = 1, \ldots, k$ for the areas. To perform the Friedman test, we assign ranks to the groups within each of the blocks and record the results in another $n \times k$ matrix $R = (r_{ij})$. The final ranking of the areas is obtained by taking an average across all blocks: $\bar{r}_{\cdot j} = \frac{1}{n} \sum_{i=1}^{n} r_{ij}$. The test statistic is obtained by

$$Q = \frac{12n}{k(k+1)} \sum_{j=1}^{k} \left(\bar{r}_{\cdot j} - \frac{k+1}{2} \right)^2 . \tag{3.1}$$

Finally, we can look up the p-value from a Friedman test table (Martin et al., 1993) and reject the null hypothesis if it is less than the predetermined significance level. If n and k are sufficiently large ($n > 15$, $k > 4$), a chi-square table can be used as an approximation. So Q is approximately χ^2_ν where $\nu = k - 1$.

3.2.2 Wilcoxon test

If the Friedman test suggests significance, we want to find out which pairs of areas are significantly different from each other. For the same reasons as in the Friedman test, we use a non-parametric hypothesis test. There are two types of Wilcoxon tests: Wilcoxon rank-sum test and Wilcoxon signed-rank test. The

Wilcoxon rank-sum test is also called the Mann–Whitney U test. This is for when the two areas are independent. However, when the areas can be paired and the measurements are dependent or related, we want to use the Wilcoxon signed-rank test (Rosner et al., 2006). For our data, even though the areas are assumed independent, the block design introduces dependency across the areas within the same block. We therefore choose Wilcoxon signed-rank test for two dependent or related samples to test for significant differences between two areas.

Consider the same data setup as above in the Friedman test with n blocks and k areas. For two areas, a, b where $a, b = 1, \ldots, k$ and $a \neq b$, the hypotheses of the Wilcoxon signed-rank test is:

$$H_0 : \lambda_a = \lambda_b$$
$$H_1 : \lambda_a \neq \lambda_b$$

We perform the test by calculating the difference between the values in the i^{th} block, denoted by $d_i = \hat{\lambda}_{ia} - \hat{\lambda}_{ib}$ for $i = 1, \ldots, n$. The absolute values of the differences are ranked. W^+ is the sum of ranks where $d_i > 0$, and W^- is the sum of ranks where $d_i < 0$. The test statistic $T = \min(W^+, W^-)$. The P-value can then be determined by comparing the test statistic with the distribution under the null hypothesis. If it is less than the predetermined significance level, then we reject the null hypothesis and conclude that there is a significant difference between the two areas.

3.2.3 False discovery rate

Moreover, if the results of many paired tests are to be considered simultaneously, we need to adjust the p-value by using the false discovery rate. In this section, we will look at how to perform these tests.

Typically in hypothesis testing, we compare the resulting p-value derived from the test statistic with the pre-set significance level (for example, $\alpha = 0.05$). This level α represents the acceptable probability of falsely rejecting the null hypothesis (or false positive rate, or type I error). This works when we are performing a single hypothesis test. However, under multiple tests, if we use the same α threshold for each of the tests, we will end up accepting too many false positives and significantly weaken the overall test. This is called the multiple comparison problem.

A method called False Discovery Rate addresses this issue (Benjamini and Hochberg, 1995). Suppose there are a total of m tests performed. We sort the p-values such that $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$. Let k be the largest i for which $p_{(i)} \leq \frac{i\alpha}{m}$. The corresponding hypotheses to the p-values that are less than $p_{(k)}$ are rejected.

3.3 Bayesian spatial models

Disease mapping is a study of disease incidence and prevalence by estimating the number of cases within geographical areas. When modelling count data, it is common to use Poisson regression models with a log-link function, so we have a likelihood $y_i \sim \text{Poisson}(\lambda_i)$, and the predictor $\eta_i = ln(\lambda_i)$ (Blangiardo and Cameletti, 2015). In addition to the log-link function, it is often needed to also include an offset, E. As the areas we are modelling on have different population sizes, using the population size as the offset to ensure that the rate is comparable across different areas. A model following equation (2.6) would then become:

$$\eta_i = \ln(\lambda_i) = \beta_0 + \sum_{j=1}^{n_1} \beta_j z_{ji} + \sum_{k=1}^{n_2} f_k(c_{ki}) + \ln(E_i) \tag{3.2}$$

where λ_i is the expected number of cases, and E_i is the population size in area i as the offset. We have $\beta_j z_{ji}$ as fixed effects that apply across all areas, and $f_k(c_{ki})$ as random effects that vary across the different areas and capture the variability within and between them.

When working with geographical data, it is often logical to think that the disease rate in neighbouring areas should be similar to each other. One natural way of accounting for this spatial effect is to build a hierarchical model where the latent field component captures the effects between the different areas. In order to extend the general hierarchical model to a spatial model, we need to know the structure of the areas in terms of their neighbours. We do so by setting up a neighbourhood map (see appendix A.2). This allows the model to capture the spatial random effects with the non-linear terms, f_k, in equation (3.2). To not lose sight of the spatial term in the hierarchical model, we express it separately and call it b_i, where i represents the areas. Hence, the model can be written as:

$$\ln(\lambda_i) = \beta_0 + \sum_{j=1}^{n_1} \beta_j z_{ji} + b_i + \ln(E_i). \tag{3.3}$$

In the following sections, we look at some commonly used spatial models for this random effect b_i.

3.3.1 The Besag model

A commonly used spatial model is the Besag model (Besag, 1974). It uses the conditional auto-regressive model (CAR) as the choice prior to model for the structured spatial effect. It is Gaussian distributed and can be written as:

$$b_i | b_{-i}, \tau_b \sim N(\frac{1}{n_{\delta_i}} \sum_{j \sim i} b_j, \frac{1}{n_{\delta_i} \tau_b}),$$

where b_i, as seen in equation (3.3), is the spatial random effect of area i and b_{-i} is the collection of all elements of b except for b_i. τ_b is the precision parameter ($\tau_b = \frac{1}{\sigma_b^2}$), $j \sim i$ denotes all j that are immediate neighbours of i, δ_i is the collection of all neighbours of area i, and n_{δ_i} is the number of neighbours. The distribution can also be read as that the mean of b_i is the average of all of its neighbours, and the precision of b_i is proportional to its number of neighbours (Moraga, 2019). Therefore the more neighbours an area has, the higher the precision which implies the lower the variance. So the field would be more smoothed out. Since it is Gaussian distributed, the joint distribution of b is therefore

$$\pi(b|\tau_b) \propto \exp\left(-\frac{\tau_b}{2}\sum_{j\sim i}(b_i - b_j)^2\right) \propto \exp\left(-\frac{\tau_b}{2}b^T Q b\right),$$

where Q is the precision matrix with the values

$$Q_{ij} = \begin{cases} n_{\delta_i} & , i = j \\ -1 & , i \sim j \\ 0 & , \text{otherwise.} \end{cases}$$

3.3.2 The Besag-York-Mollié (BYM) model

The Besag model assumes only the structured spatial effect based on the neighbourhood map. Any variability in the regions that is not due to the geographical placement will be forced into the structured effect. This may not be desirable. An improvement from Besag would be the Besag-York-Mollié (BYM) model (Besag et al., 1991). It splits the spatial effect into structured and unstructured components, $b = u + v$. Here, u is the structured effect captured by the Besag model with $u \sim N(0, \tau_u^{-1}\mathrm{Q}^-)$ where Q^- is the generalized inverse of Q (where $\mathrm{QQ}^-\mathrm{Q} = \mathrm{Q}$), and v is the unstructured effect with $v \sim N(0, \tau_v^{-1}\mathrm{I})$. So the covariance matrix is

$$\mathrm{Var}(b|\tau_u, \tau_v) = \tau_u^{-1}\mathrm{Q}^- + \tau_v^{-1}\mathrm{I}. \tag{3.4}$$

3.3.3 The Leroux model and the Dean model

Even though the BYM model is an improvement from the Besag model, the two spatial components still cannot be seen independently. In addition, τ_u and τ_v do not represent variability on the same level. This makes the choice of hyperpriors very challenging. Some modifications, such as the Leroux model and the Dean model, have been proposed to address this problem.

The Leroux model introduces a mixing parameter, $\phi \in [0, 1]$, to capture the balance between the two components (Leroux et al., 2000). More specifically, instead of the covariance matrix in equation (3.4), b

now has the covariance matrix

$$\text{Var}(b|\tau_b, \phi) = \tau_b^{-1}(\phi \text{Q} + (1-\phi)\text{I})^{-1}. \tag{3.5}$$

Not only does this give a simplification for the hyperpriors, it also allows us to interpret the strength of the spatial component with the hyperparameter ϕ. When $\phi = 0$ the model suggests no structured spatial effect and when $\phi = 1$ then the model is equivalent to the Besag model.

The Dean model proposes a reparameterization of the BYM model with $\tau_u^{-1} = \tau_b^{-1}\phi$ and $\tau_v^{-1} = \tau_b^{-1}(1-\phi)$ (Dean et al., 2001). This gives us

$$b = \frac{1}{\sqrt{\tau_b}}(\sqrt{1-\phi}\, v + \sqrt{\phi}\, u)$$

with the covariance matrix

$$\text{Var}(b|\tau_b, \phi) = \tau_b^{-1}(\phi \text{Q}^- + (1-\phi)\text{I}).$$

3.3.4 The BYM2 model

A further evolution of the BYM model, called the BYM2 model, proposes scaling the structured component, u, in the Dean model so that the hyperpriors can be used and compared across multiple models (Riebler et al., 2016). The generalized variance of u can be expressed as

$$\sigma^2_{GV}(u) = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{1}{\tau_b}[\text{Q}^-]_{ii}\right)\right) = \frac{1}{\tau_b}\exp\left(\frac{1}{n}\sum_{i=1}^{n}\log\left([\text{Q}^-]_{ii}\right)\right).$$

To scale it we want $\sigma^2_{GV}(u) = \frac{1}{\tau_b}$, which can be interpreted as that τ_b represents the precision of the marginal deviation from a constant level, regardless of the spatial structure. Let the scaled u be u_*, and the corresponding precision matrix be Q_*, then we have

$$b = \frac{1}{\sqrt{\tau_b}}(\sqrt{1-\phi}\, v + \sqrt{\phi}\, u_*) \tag{3.6}$$

with the covariance matrix

$$\text{Var}(b|\tau_b, \phi) = \tau_b^{-1}(\phi \text{Q}_*^- + (1-\phi)\text{I}).$$

BYM2 model remains to have the intuitive interpretation of ϕ being the proportion of the marginal variance explained by the structured spatial effect. Moreover, the scaled structured spatial effect u_* results in a scaled random effect b and standardized interpretation of the hyperprior τ_b across different models.

To apply the PC prior, we specify U_1 and α_1 for the precision parameter τ_b such that $P(1/\sqrt{\tau_b} > U_1) = \alpha_1$. Note that the larger the U_1 the more variability we are allowing in the distribution which results in a more non-informative prior and allowing the data to dominate the model. Similarly, the prior for the mixing

parameter ϕ, which controls the proportion of the marginal variance of the structured spatial effect u_*, can be built by specifying U_2 and α_2 such that $P(\phi < U_2) = \alpha_2$.

3.4 Analysis on geographical variation in stroke incidence rates

In this section we perform 2 types of analysis. First we use hypothesis test within the frequentist framework to see if there are any significant differences in the stroke incidence rates between the geographical areas. If so, we investigate which areas are different. This is repeated at two different levels - starting off with a high level view of 4 regional health authorities and progressing to a more detailed level with 21 hospital referral areas. The other type of analysis is by using spatial models within the Bayesian framework. Instead of testing and finding significance, we use a model to capture the underlying pattern in the data. We explore models with different components to see which ones describe the data best. A model comparison is done at the end of this section as a summary for the final model chosen.

3.4.1 Hypothesis tests in frequentist framework

Here we test if the incidence rates between the geographical areas are significantly different with an α level of 0.01. As shown in the data section in figure 2, the age and gender make-up in each area is different and affects the incidence rate. We therefore want to test with a block design. Each patient is assigned to one of the 11 age blocks as defined in section 3.1. Moreover, the genders are also kept in separate blocks, thus 22 age and gender blocks all together in each geographic area. This allows the rates to be comparable across the areas.

Regional Health Authorities (RHA)

Friedman test In the data section, we see the adjusted rates for the 4 RHAs differ slightly (figure 3). Here we would like to perform hypothesis tests to see if they are significantly different. Since the underlying data is count data (number of patients) and we have the age and gender blocks, we use a non-parametric Friedman test (see section 3.2 for more details). The null hypothesis is that $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$. The result of the test gives us a p-value of $2.82 * 10^{-6}$, which is much less than α, so we reject the null hypothesis and conclude that the stroke incidence rates for the regions are significantly different.

Wilcoxon test To further study the difference, we perform paired tests to find out which of the regions differ from each other. We use Wilcoxon signed-rank test for the count data. In addition, since there are multiple tests, we need to adjust the resulting p-value with false discovery rate. Table 1 shows the adjusted

p-value. There are no significant differences between the north and the central regions, as well as between the west and south-east regions, with p-values of 0.1181 and 0.3880 respectively. There is however difference between the two regions that are above the national rate (north and central) and the other two (west and south-east). This result agrees with what we saw in figure 3.

RHA	North	Central	West	South-east
North		0.1181	0.0016	0.0016
Central			0.0011	0.0006
West				0.3880
South-east				

Table 1: P-values from the Wilcoxon signed-rank test between the regional health authorities adjusted with false discovery rate.

Hospital referral areas (RA)

The 4 regions each cover very large geographical areas and could have much variation within. We are interested in seeing if the difference is observed at a finer level, namely the hospital referral areas. Appendix A.1 shows the referral areas that belong to each of the regions. Since the Friedman test at the regional level shows significant differences, we can safely assume that the differences remain at a finer level and omit performing the Friedman test. We go directly to the Wilcoxon signed-rank test and adjust the results with false discovery rate to find out which pairs of RAs have significantly different incidence rates from each other.

Table 2 shows the 21 RAs down the left, as well as across the top. The RAs that belong to the same RHA are grouped together with lines between the different RHAs. The crosses in the table indicate the pairs with incidence rates significantly different from each other based on the Wilcoxon test.

Referral area Diakonhjemmet (RA 16) has the lowest incidence rate (see figure 3) but is statistically significant from only 7 of the other 20 RAs. More specifically, it is not different from any of the 9 RAs that are also below the national average. Referral area OUS (RA 14) has the highest incidence rate and is statistically significant from 13 of the other 20 RAs, and not from 7 of the top areas. This indicates that while we see differences, there are not that many, and not that pronounced. If we look at the RHAs, we can see that all RAs within the north and central RHAs are not different from each other. Instead, they are more different from the other regions, especially areas in south-east RHA. This suggests that there may be some spatial pattern separating the northern and central regions from the rest of the country.

RHA	North				Central			West				South-east									
RA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. Finnmark																					
2. UNN									X	X	X	X	X		X	X		X	X	X	X
3. Nordland																			X		
4. Helgeland																			X		
5. Nord-Trøndelag																					
6. St. Olavs									X				X			X		X	X	X	X
7. More-Romsdal									X				X	X		X		X	X		X
8. Førde														X		X					
9. Bergen		X				X	X							X			X				
10. Fonna		X												X							
11. Stavanger		X												X							
12. Østfold		X												X		X					
13. Akershus		X				X	X							X							
14. OUS							X	X	X	X	X	X	X		X	X		X	X	X	X
15. Lovisenberg		X												X							
16. Diakonhjemmet		X				X	X	X				X		X			X				
17. Innlandet									X							X			X		X
18. Vestre Viken		X				X	X							X							
19. Vestfold		X	X	X		X	X							X			X				
20. Telemark		X				X								X							
21. Sørlandet		X				X	X							X			X				

Table 2: 'X' marks where there is difference in the incidence rate at the 0.01 significance level between the corresponding hospital referral areas.

3.4.2 Spatial models in Bayesian framework

In this section, we investigate stroke incidence rates with models within the Bayesian framework. As the data is the number of stroke patients in the different geographic areas in Norway, it is natural that we explore with a spatial model to identify if there are any underlying spatial patterns.

As we saw from figure 1, the older the age group, the higher the stroke incidence rate. In addition, men are generally more likely to get stroke than women. Just as when we calculate the age and gender adjusted incidence rate in section 3.1, we need to take these differences into consideration and structure our model to reflect this.

We set up a joint model (as described in section 2.2.5) with two response vectors - one for men and one for women. Each of them will have their own intercept and age blocks in order to capture the variation we

observed by age. However, they will have a shared random effect, namely place of residence (hospital referral area), since regardless of age and gender, the patients that live in the same area are expected to be impacted similarly. It is this spatial effect we are focusing on when doing a geographical study.

We will approach the model building process in a step-by-step fashion. First we start with an empty model where we have just the responses and the intercepts. This gives us an idea of what the baseline is. Then we include the age blocks as random effects to capture the patterns in the age groups. Lastly we explore the spatial effects by using IID and BYM2 as the latent model component.

The models will be examined with the following output:

1. Posterior mean of the modelled rate, which is calculated by transforming the linear predictor with the inverse of the link function, then multiplied by the offset. This is done for each of the age and gender grouping within each RA, and then aggregated to get a total number of expected cases per 10 000 population per RA. This is visualized on a plot with the actual number of cases (also age and gender adjusted) as a comparison.

2. Posterior mean and 95% credible interval (CI) for the random effect from the age blocks. If the CI for the blocks from male and female overlap, then that is an indication that there is no significant difference between the genders within the blocks. However if they don't, then joint model with two responses enables us to explain the gender differences separately.

3. Posterior mean and 95% credible interval for the random effect from the RAs. There are a few things we can infer from this. First of we examine the length of the CIs, if they are quite long then that means there is much uncertainty. Secondly we see if there is overlap of the CIs between the RAs. Just as for the age blocks, if the CIs overlap, then they are not significantly different from each other, and vise versa. Lastly, if the CI covers 0, then there is no significant spatial effect from the RA.

Model with intercepts

Here we are using the Poisson distribution as the likelihood function, and the population size as the offset. We are therefore modelling the rate, λ, with logarithm as the link function. With $i = 1, \ldots, 21$ for the areas, $j = 1, \ldots, 11$ for the age blocks and $g = (m, w)$ for the genders, the model can be written as

$$\ln(\lambda_{gij}) = \beta_g + \ln(E_{gij}),$$

where β_g are the intercepts for the two genders and E_{gij} are the population size for each age and gender in each RA as the offset.

The posterior mean and 95% CI for the intercepts are -5.601(-5.632, -5.571) and -5.411(-5.439, -5.383) for women and men respectively. Notice that the CIs do not overlap, which means that the effect of men and women are significantly different from each other. This validates the need of modelling the genders separately and setting it up as a joint model.

Figure 4 illustrates the crude rate from the original data compared to the modelled rate. Along the x-axis are the RAs, with indications of the RHA groupings (RA 1-4 are in RHA north, RA 5-7 are RHA central, RA 8-11 are RHA west, and RA 12-21 are RHA south-east). The top two figures illustrate the results for women and men plotted separately. We see that women have lower crude rates and consequently result in lower intercepts. Since the intercepts are constants and the model does not depend on any other covariates, the overall modelled rate shown in the bottom figure is also constant and represent the national average. We will further explore other factors that may help explain the gaps between the crude rate and modelled rate.

Model with age blocks

As seen from the data chapter 3.1, we know that the stroke incidence rate is highly correlated with age. Therefore as a next step from the model with only intercepts, we now add the age blocks for both genders. Since figure 1 (right) suggests that there is a non-linear relationship between the age blocks and the percentage of population with stroke, we use RW2 as the latent model for the random effects. This together with the two response vectors make sure that the model is build for each age and gender group. The model then has the form:

$$\ln(\lambda_{gij}) = \beta_g + b_{gij} + \ln(E_{gij}),$$

where b_{gij} is the random effects for the age blocks j and gender g in area i.

Figure 5(top and bottom left) shows the crude and modelled patient rate for each RA with the respective average rates as the reference lines. We see that since the sizes of age blocks vary in each RA, it already gives us some guideline in bringing the modelled rates closer to the crude rates. Figure 5(bottom right) shows the age effect with the red line for the women and blue for men. The grey areas around the lines mark the 95% CI. It is worth noting that for most of the age groups, the CI do not cross, which means that men and women have significantly different age effects. However, the lines do cross at around age group 75-79. This indicates that there is a switch from men having a strong effect to women as the population age. Based on figure 1(left), we can say that this is due to the higher volume of male patients in the younger age groups

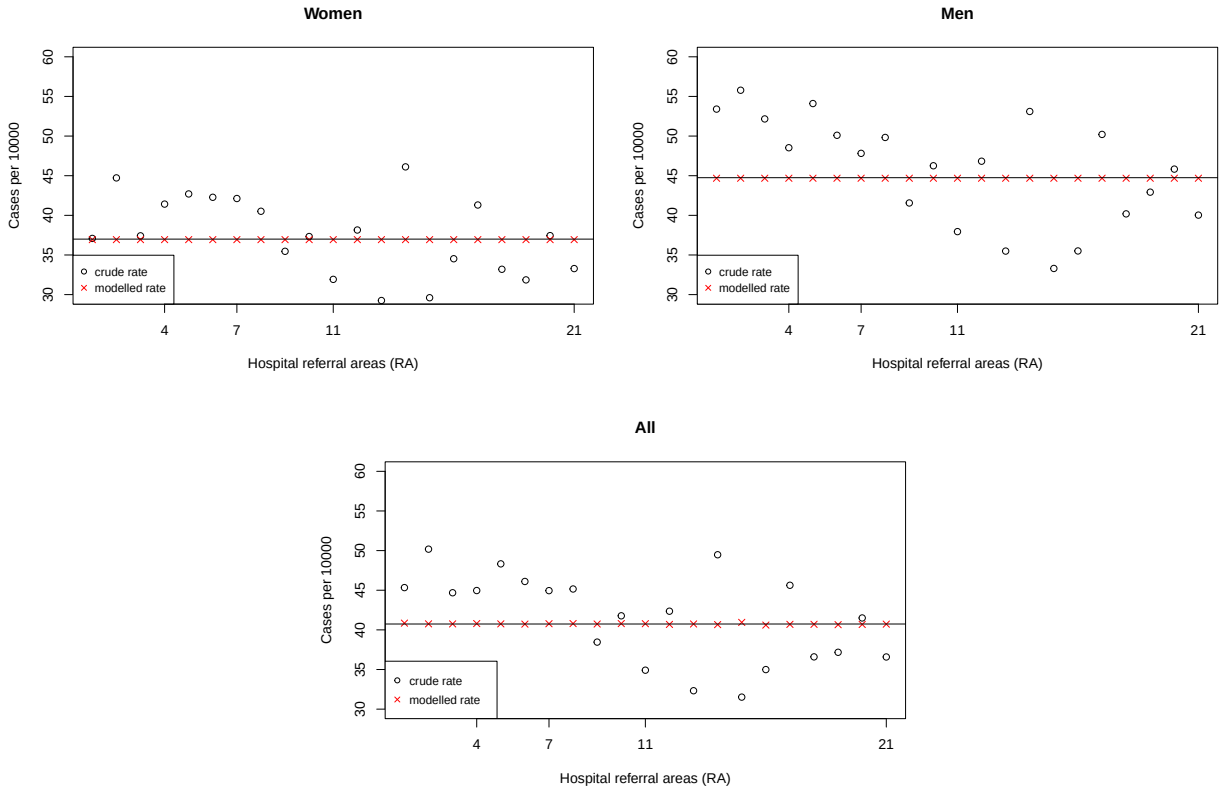

Figure 4: Model with intercepts. Crude incidence rate (black circle) compared with modelled rate (red cross). The top left figure is for women. The horizontal reference line at 37 denotes the national average for women. The top right figure is for men with the reference line at 44.75 denoting national average for men. The bottom figure is for the entire population with the reference line at 40.74.

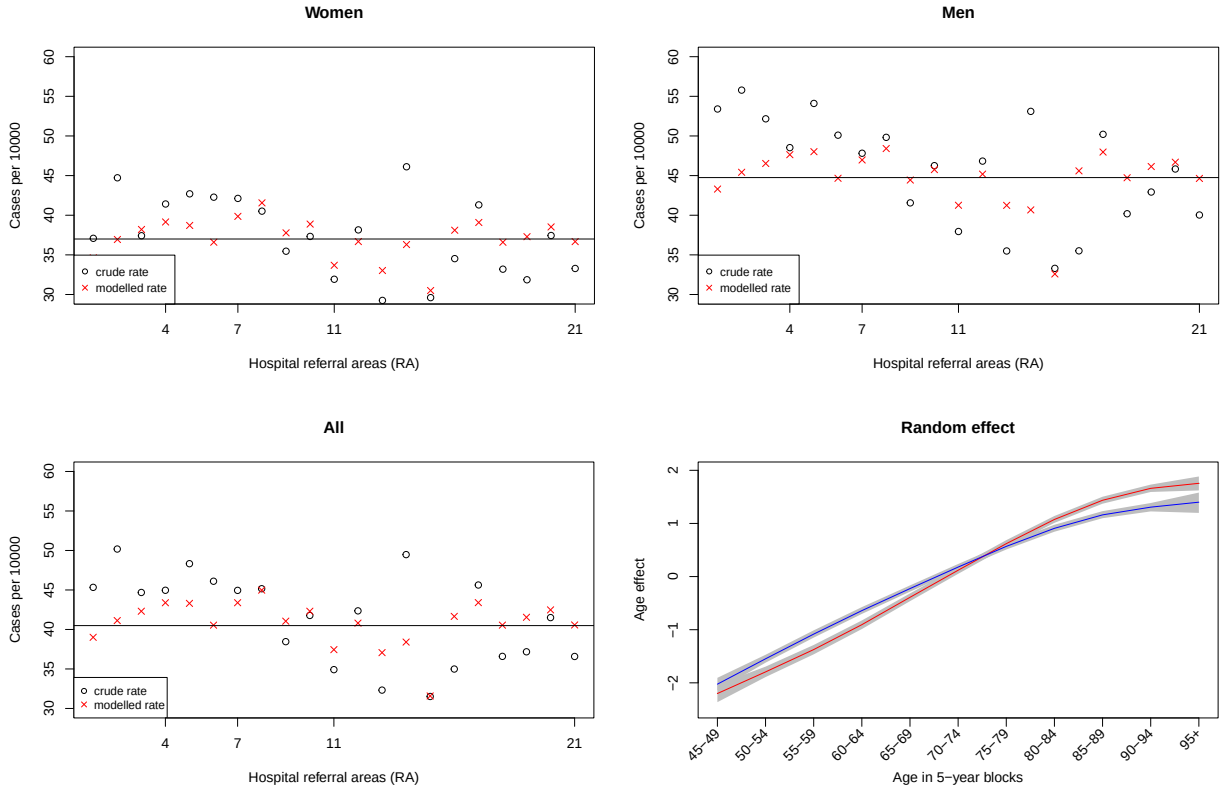

Figure 5: Model with age blocks. (Top two and bottom left) Crude incidence rate compared with modelled rate for women, men and overall respectively. (Bottom right) The age effect on the modelled rate by age block and gender. Red is for women and blue is for men.

that switched to higher volume of female patients in the older age groups. This effect is observed by another research as well (Roy-O'Reilly and McCullough, 2018).

Model with IID spatial effects

While the age effects noticeably improve the model in explaining the stroke incidence rate, there are still large gaps between the crude rates and the modelled rates that remain unexplained. We are now ready to explore if there are spatial effects to be capured. We do so by including the RAs as a random effect. However, here we want to use it as a shared term between the genders in a joint model, and first investigate unstructured spatial effect with an IID latent model. The model takes on the form:

$$\ln(\lambda_{gij}) = \beta_g + b_{gij} + u_i + \ln(E_{gij}), \tag{3.7}$$

with an additional term u_i representing the RAs as a shared effect between the genders following an IID model. We use PC prior for the hyperparameter of u_i with $U = 1$ and $\alpha = 0.01$.

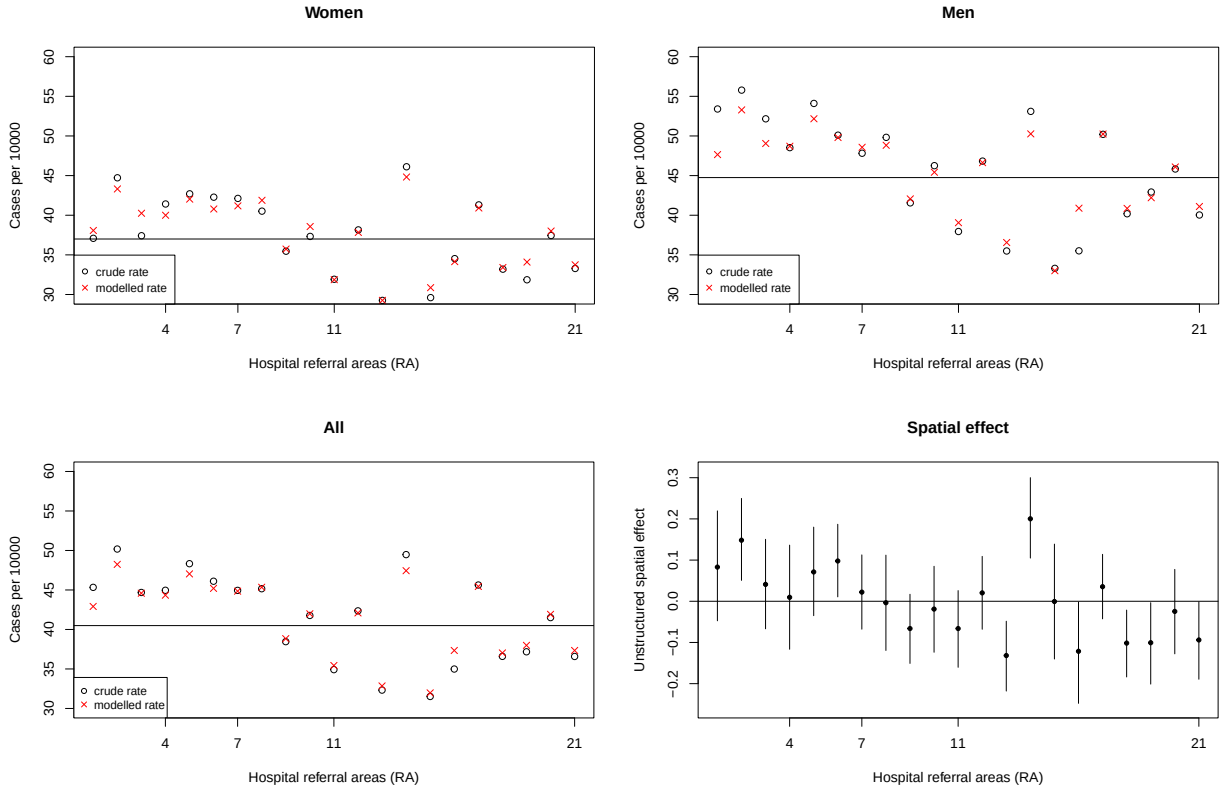

Figure 6: Model with IID latent component for the RAs. (Top two and bottom left) Crude incidence rate compared with modelled rate for women, men, and overall respectively. (Bottom right) The unstructured spatial effect (IID) of the RAs. The dots are the posterior means and the lines mark the 95% CI.

Figure 6 shows that the modelled rate is much closer to the crude rate compared to the previous two models. The age random effect is still present, but remains as before, therefore not shown on the figure. On the bottom right of figure 6, we see the random effects of the RAs captured by the IID model. The dots represent the posterior means while the lines span the 95% CI. Even though the means seem to vary between the RAs, many of them overlap and most of them have a CI that includes 0. This means that there are no unstructured spatial effect for most RAs. Some exceptions are OUS (RA 14), UNN (RA 2) and St. Olavs (RA 6) with positive effects, and Akershus (RA 13) and Vestre Viken (RA 18) with negative effects.

Model with BYM2 spatial effects

So far we see that there are age effects as well as unstructured spatial effects in a few RAs on stroke incidence rates. In addition to these, we want to see if there are any structured spatial effect that takes the geographical structure into consideration by using the neighbourhood map (see appendix A.2). We have

chosen to use the BYM2 model here to include both the unstructured and structured spatial effects. The formulation of the model is the same as in equation (3.7), but with u_i being a BYM2 model instead of IID. We use a PC prior and specify for the two hyperparameters: the precision parameter τ and the mixing parameter ϕ. After a sensitivity analysis, we ended up with $U = 1$, $\alpha = 0.01$ for τ, and $U = 0.5$, $\alpha = 0.5$ for ϕ.

We can see from figure 7 that there are only minor shifts in the modelled rates from the BYM2 model compared to the model with IID effect (figure 6). The age random effect remains the same and there are only minor shifts in the unstructured spatial effect, hence we do not repeat this plot and instead focus on the structured spatial effect. Bottom right of figure 7 shows the spatial effect captured when using the neighbourhood structure. The means suggest that RHA north and central (RA 1-7) have positive effects while RHA west and south-east (RA 8-21) have negative effects. However, if we look at the 95% CI, we see that there are a lot of overlap and all of the areas have a CI that covers 0. This means that there are no structured spatial effect. Furthermore, the mixing parameter ϕ has posterior mean and 95% CI of 0.29 (0.03, 0.74). This indicates a weak spatial relationship.

3.4.3 Model selection

We have seen a progression of models and can now compare the models to choose the one that describes the data best. While there are many different metrics that can be used to compare Bayesian models, we choose to use DIC as described in section 2.2.6. Table 3 below shows that DIC for the intercept model is high as expected. It dropped drastically for the model with age blocks, and even more with the IID spatial component. However, introducing the structured spatial effect in BYM2 did not change the measure. The BYM2 model introduces complexity without much improvement. Therefore we choose the IID model as our final model and conclude that there are no structured spatial effects in stroke incidence rate.

Model	DIC
Intercept	11944
Age block	2287
IID	2215
BYM2	2215

Table 3: DIC values for the models

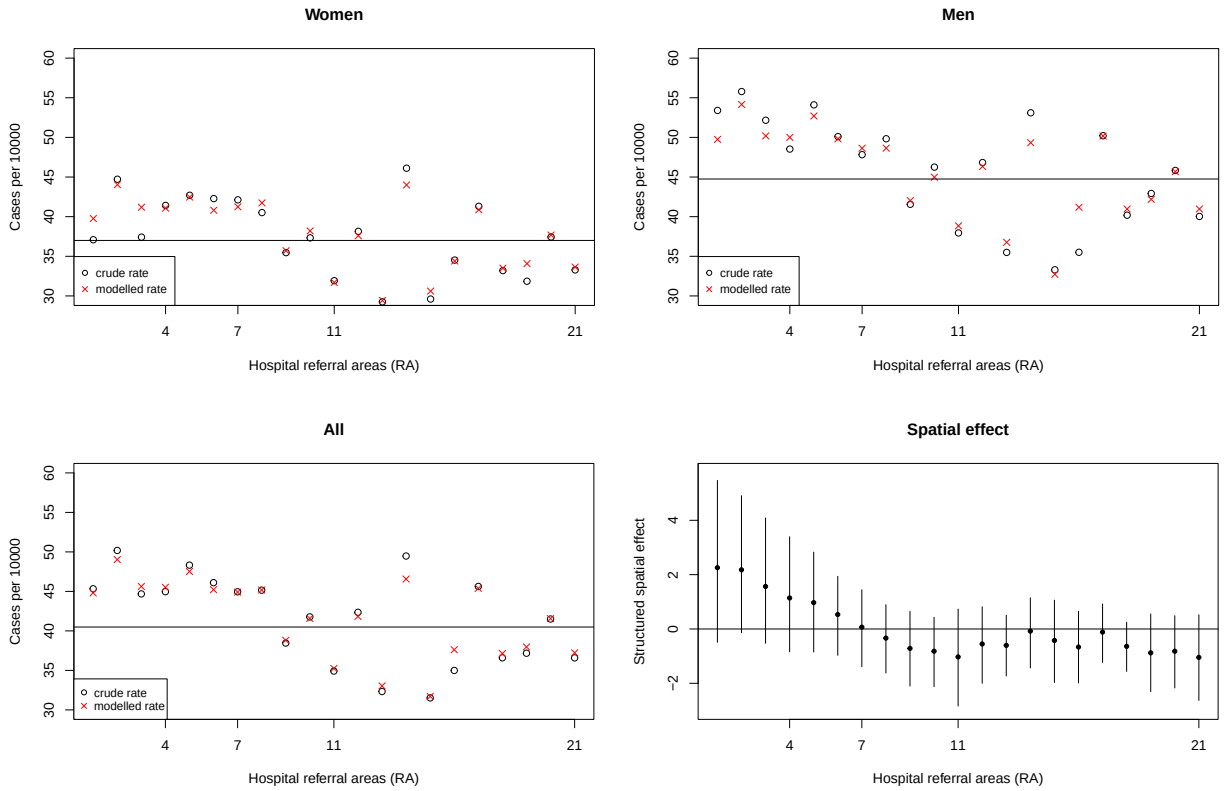

Figure 7: Model with BYM2 latent component for the RAs. (Top two and bottom left) Crude incidence rate compared with modelled rate for women, men, and overall repectively. (Bottom right) The structured spatial effect (BYM2) of the RAs. The dots are the posterior means and the lines mark the 95% CI.

3.5 Remarks on geographical variation in stroke incidence rates

In this chapter we have examined the geographical variation in stroke incidence within Norway using hypothesis tests from the frequentist framework and spatial models from the Bayesian framework.

The hypothesis tests show that there are no significant differences in the stroke incidence rate within the north and the central regional health authorities. However these two regions are significantly different from the other two regions in the country. Moreover, when we look at the country in a slightly finer geographical detail with the hospital referral areas, the hypothesis tests suggest that most areas have incidence rates significantly different from at least one other area. Of which the most noticeable are OUS and UNN which are the two RAs with the highest incidence rates (table 2). On the contrary, the spatial models indicate that there are no structured spatial effects (figure 7). When it comes to the unstructured spatial effects, most areas have no significant effects, aside from OUS, UNN and St. Olavs with positive effects and Akershus and Vestre Viken with negative effects (figure 6).

One may observe this to be contradictory results from the the different methods. However, they are not directly comparable. The hypothesis tests assume independence between the areas, which we know is not necessarily true. The people that live in each area as well as the environment they live in and the lifestyle they have can very well be related to the other areas. The violation of the assumption may result in an underestimation of the variance hence giving us shorter confidence intervals and therefore more likely for significance. The spatial model on the other hand makes no assumptions and allows the model to capture spatial effects, if any. Therefore we argue that the results from the spatial model reflect the data better.

While some of the details of the results from the hypothesis test and the spatial model differ, they both indicate that OUS and UNN stand out as the two RAs with significantly higher incidence rates than the other areas. We therefore conclude that aside from these two RAs, there are no significant geographical variation in stroke incidence rate.

Norway is a very widespread country and dividing it into only 21 geographical areas is potentially too coarse. Particularly when it comes to the structured spatial effect. Not only are there too few areas, since Norway is a long country, the majority of the areas don't have many neighbours. This means that we are not taking full advantage of the neighbourhood structure and that some of the area specific details may be hidden in the data. A logical next step is to subdivide the referral areas into even finer scale, for example at the municipality level, in order to capture potential pockets that may have different patterns.

4 Survival analysis

In the previous chapter, we looked at the incidence rate of stroke patients and explored if there are significant spatial patterns to explain the geographical variation. In this chapter, we will further look at these patients and study their survival rates. In particular, we aim to examine different factors to see if they contribute to the survival rate. These factors include demographics, socioeconomic status, and geographical location.

We will first investigate the data in the aforementioned categories in section 4.1. Section 4.2 outlines the theory behind survival analysis including the survival and hazard function, censoring, Kaplan-Meier analysis, and Cox proportional-hazard models. Section 4.3 provides some background on the implementation of the survival model in INLA. This is then followed by section 4.4 where we perform the Kaplan-Meier and Cox proportional hazards (Cox PH) analysis on stroke patients and see which factors affect the survival. Finally section 4.5 summarizes the results.

4.1 Mortality rates of stroke patients

For the survival analysis, we use the same patient population as described in section 3.1. Patients from the Norwegian stroke registry (NHR) during 2014-2018 with no previous stroke incidence in the same period are identified as stroke patients. These patients are subsequently matched with the Norwegian cause of death registry (DÅR). The data from DÅR is from 2014-2020 so that each patient is followed for at least 2 years before the end of the study period. We therefore have the information on which of these stroke patients did not survive during the period, date of death, and cause of death. For the purposes of this thesis, we have chosen to identify cause specific stroke mortality by using main cause as well as secondary causes. Those that died within the study period and have the cause of death diagnosis codes I61, I63, I64 [4] as one of the causes are identified as "stroke death", and those that do not have stroke listed as one of the causes are therefore identified as "other death". Figure 8 shows the mortality rate for each of the two groups by number of weeks after the stroke incidence.

[4]ICD-10 code I61 is for Nontraumatic intracerebral hemorrhage, I63 is for cerebral infarction, and I64 is for stroke, not specified as hemorrhage or infarction.

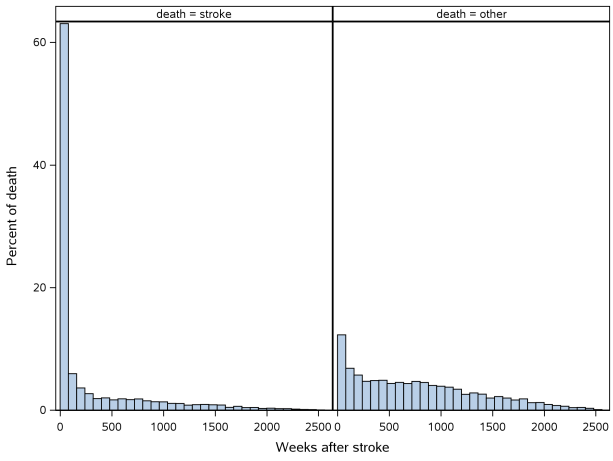

Figure 8: Percent of death by week for stroke deaths and other deaths

There are 30 302 stroke patients from the stroke registry. Nearly 40% died before the end of 2020, of which 5 000 are stroke specific deaths and 6 000 are others. We are interested in seeing if there are any significant factors within patient demographics, health characteristics, socioeconomic status, as well as geographical differences, that contribute to mortality.

Demographics

First we look at patient demographics by age and gender. Age groups are set up the same way as in the spatial model: 11 groups starting at the age of 45 with 5 years per group, and the last group being those of age 95 and older. As expected, the mortality rate is higher for older stroke patients (shown in figure 9 (left)). We can also see that men and women have different mortality rate (figure 9 (right)) with women having the higher rate. However, we know that there are more women in the higher age group (figure 1) and consequently the stroke patient population would have more older female patients than older male patients. Furthermore, we know that older patients have a higher mortality rate. These in combination can contribute to the result of a higher mortality rate for women when we use the entire stroke patient population.

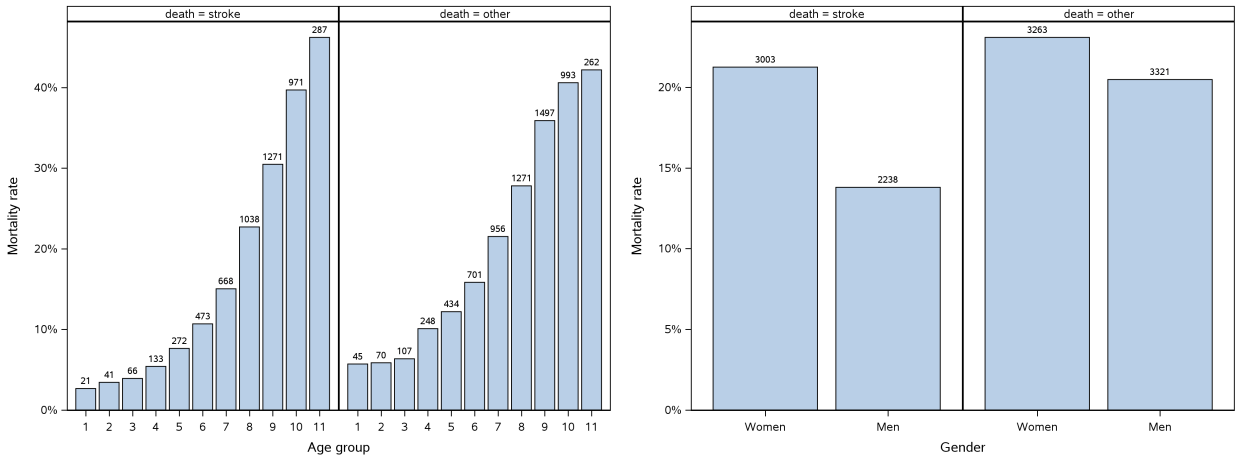

Figure 9: Mortality rate by patient demographics age (left) and gender (right)

In order to have a more comparable population, we create stratified samples of 30 patients from each age group within each gender [5]. This way the effect of age is removed and we can examine the mortality rates between genders more clearly. In figure 10, we see that men between ages 70-85 (age group 6-8) have a higher stroke mortality rate than women in the same age group. On the other hand, women younger than 70 or older than 85 have a higher mortality rate than their male counter-parts. However, for mortality caused by reasons other than stroke, men have a higher rate in almost all age groups. This is especially pronounced for 85 years and older.

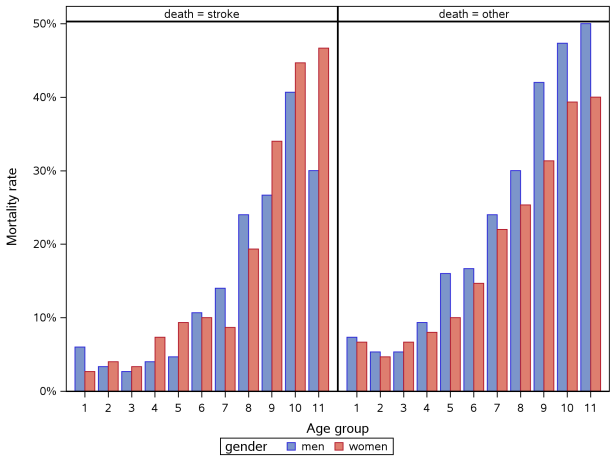

Figure 10: Mortality rate by patient demographics in stratified sample - gender within age group

[5] We have chosen to sample 30 due to the limitation of some of the older age groups where there are less patients. According to the central limit theorem, this is a large enough sample to represent the population.

Next we look at the type of stroke - hemorrhagic (bleeding) or ischemic (blockage), as well as time spent in the hospital, see figure 11. Hemorrhagic stroke is often a more serious condition than ischemic stroke and therefore has a higher mortality rate. Time spent in the hospital is calculated by number of days between arrival to the hospital and discharge or death. It is categorized [6] for the purpose of data exploration. Since stroke is an acute condition and many die within a short time after the onset of the condition, there appears to be quite a high mortality rate for those that spent less than 3 days in the hospital, i.e. category 1. Previous studies have shown that about half of the death due to hemorrhage usually occurs within the first three days (Fernando et al., 2021; Rymer, 2011). After the initial period, the rate seems to drop as it can indicate that the patients who survived up to that point have decreased risk. However, as a patient stays longer in the hospital, their conditions may have worsen and results in a higher mortality rate. Due to this non-monotonic relationship, we will be using a random walk 2 model (see section 2.2.3) to capture the non-linear pattern.

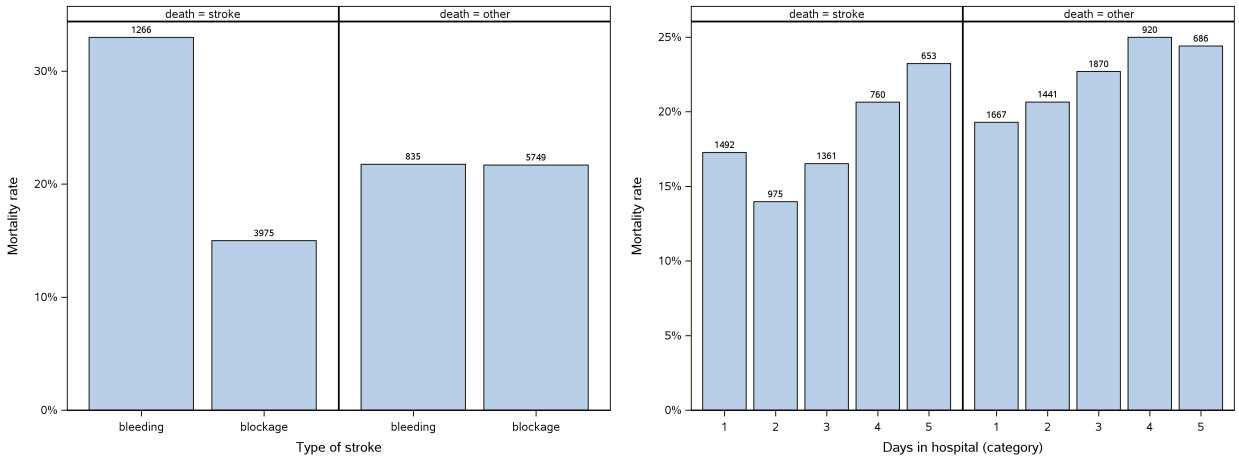

Figure 11: Mortality rate by patient health characteristics: type of stroke and time spent in hospital

Socioeconomic

For the socioeconomic status, we have income and education levels from Statistics Norway for each individual. Education level is put into 3 categories [7]. From figure 12 (left), we see that the long-term mortality rate decreases with higher education. Income level is determined by using after tax household income, in Norwegian kroner (NOK). Since we have income for a span of 5 years, in order to take into account inflation

[6] Time spent in hospital categories: 1 = 0-3 days, 2 = 4-5 days, 3 = 6-9 days, 4 = 10-14 days and 5 = more than 14 days.
[7] Education levels: low = grades 1-10 (obligatory), medium = up to high school, high = bachelors and above.

and make the numbers more comparable, we have adjusted with the consumer index [8]. It is then put into 4 categories [9]. In figure 12 (right), we see that there is a similar trend as education where patients with higher income have lower mortality rate.

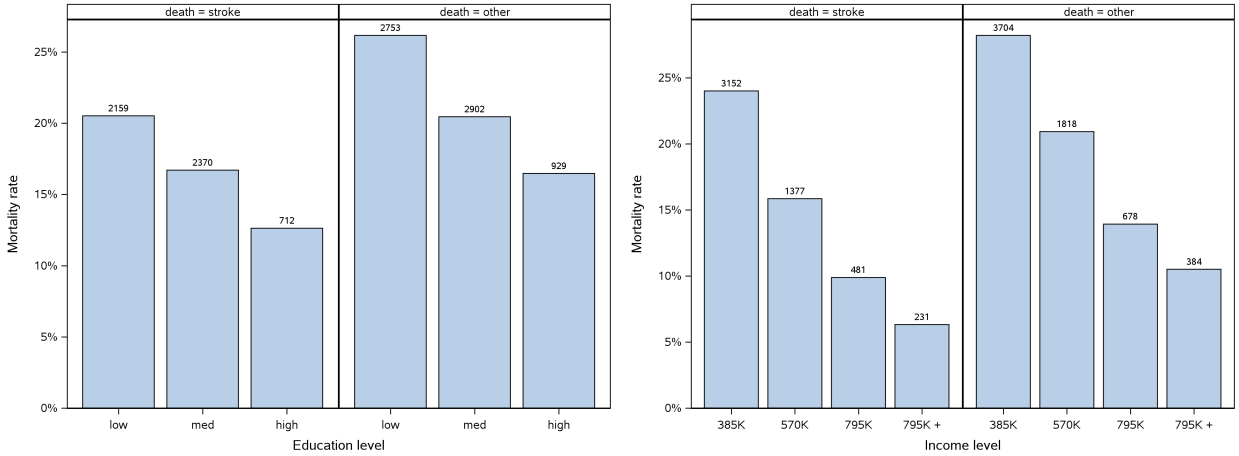

Figure 12: Mortality rate by patients' socio-economic status: education and income

Geographic

Finally, we examine two variables to see the effect of geographical differences: place of residence and travel time to the local hospital. The place of residence is the hospital referral area (RA) where the patient lives at the time of stroke incidence. RA is defined the same way as in the spatial model (appendix A.2). According to figure 13, there does not appear to be any clear relationship between where one lives and mortality rate. Travel time is the number of minutes it takes by car from the center of the municipality the patient lives into the nearest local hospital with acute care function. This is categorized into 5 groups [10]. There seems to be an increase in mortality rate for those that live between 0-90 minutes from the hospital, but a decrease for those further than 90 minutes. This may seem a bit counter-intuitive. One possible explanation is that since stroke is an acute condition, it is crucial that patients receive care as soon as possible. Therefore the longer it takes to reach the hospital the higher risk of death. However, those who live further away are likely not transported by car but instead by air, hence reaching the hospital faster which increases survival.

[8] Norwegian consumer price index from Statistics Norway `https://www.ssb.no/en/priser-og-prisindekser/konsumpriser/statistikk/konsumprisindeksen`

[9] Income levels: up to 385 000 NOK, 385 001 - 570 000 NOK, 570 001 - 795 000 NOK, and lastly 795 001 and above.

[10] Travel time categories: 1 = 0-30 minutes, 2 = 31-60 minutes, 3 = 61-90 minutes, 4 = 91-120 minutes and 5 = more than 120 minutes.

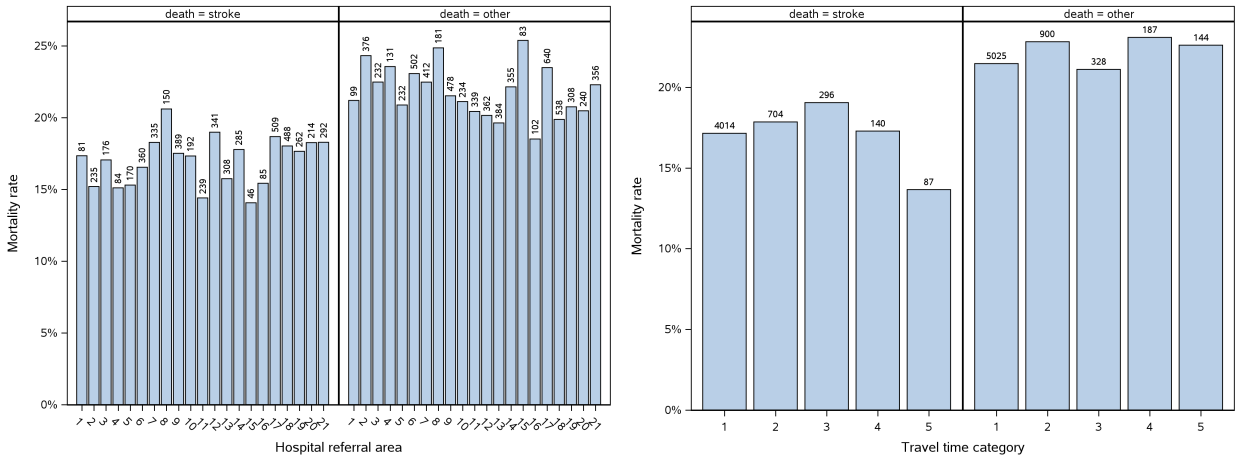

Figure 13: Mortality rate by geographic: place of residence and travel time to hospital

In this section we have looked at one-dimensional data profile to get a better understanding of the data. We have provided intuitive interpretations to the observed pattern in some cases. However, these do not show the true effect of the potential risk factors. Often there are inter-relationships to the factors that would change the significance of the final results. The survival analysis later on in section 4.4 will show if the propositions are reasonable and if the effects are considered significant when all factors are modelled together.

4.2 Theory on survival estimator and model

The main interest in survival models is the time to event. Let T be a random variable that measures time to event and $f(t)$ be the probability density function of T. The cumulative distribution function is then the probability of an event occurring prior to time t and can be expressed as

$$F(t) = P(T \leq t) = \int_0^t f(u)du, \quad t > 0.$$

From here, we can then arrive at the probability of survival, which by definition is the probability of event occurring after time t. The survival function is the complement of the cumulative distribution function and can be written as

$$S(t) = P(T > t) = 1 - F(t), \quad t > 0.$$

Next we define the hazard rate as the instantaneous event rate. In other words, at time t, what is the probability of the event occurring in the next instant (Δt) given that the subject has survived up until that

point. This can be expressed mathematically as

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{P(T < t + \Delta t \,|\, T > t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t)}{\Delta t \cdot P(T > t)} \\ &= \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t \cdot S(t)} \\ &= \frac{f(t)}{S(t)}. \end{aligned}$$

Lastly, we can derive the cumulative hazard function which tells us the total number of events occurred up to time t. By using the information we have from above, we now have

$$\begin{aligned} H(t) = \int_0^t h(u)du = \int_0^t \frac{f(u)}{S(u)}du \\ &= \int_0^t \frac{\frac{d}{du}F(u)}{S(u)}du \\ &= \int_0^t \frac{\frac{d}{du}(1 - S(u))}{S(u)}du \\ &= \int_0^t \frac{-S'(u)}{S(u)}du \\ &= -\log(S(t)). \end{aligned} \tag{4.1}$$

Additionally, it can be useful sometimes to express the survival function in terms of the cumulative hazard function.

$$S(t) = \exp(-H(t))$$

4.2.1 Censoring and truncation

Survival data is often incomplete in the sense that not all subjects would have necessarily experienced the event within the study period. However, despite the potential incompleteness of the data, these observations are still informative and do affect our interpretation of the survival rate. Hence, we do not want to simply exclude them.

In order to capture the status of the data correctly, one of the methods often used is censoring. There are many different types of censoring, depending on the situation. Some examples are right censoring, left censoring, and interval censoring. To capture the censored data, 2 sets of information is required to describe the dependent variable of a survival model. For each subject i, we capture event at time t_i, as well as censoring status δ_i. δ_i is 1 when the subject experienced the event, with time of event at time t_i. δ_i is 0 when the subject was in the study population until then but exits the study at the corresponding time t_i

without experiencing the event, hence censored.

The difference between the above mentioned types of censoring is the time point at which the record is censored relative to the actual, but unobserved occurrence of the event. Right censoring is the most common type of censoring. This is when the unobserved event happens after the censored time point. This could be due to reaching the end of study period, or other reasons that cause the subject to exit before the end of the study period. For example, consider a study of cancer patient survival for a follow up time of 2 years. At the onset of the study, all patients are alive. Those that did not die by the end of the 2 years are censored, as well as those that emigrated and therefore can not be followed. In contrast, left censoring is for when the event happened before the censored time point, and interval censoring is when the event happened between two time points.

In addition to censoring, which is done at the individual record level, there may be situations when certain records are excluded from the data set. This is called truncation. When time-to-event is too short and the data collection process or apparatus is unable to include the subject in the data, it is a case of left truncation. However, a right truncation is when time-to-event is too long and the subject is excluded due to the limitation of the study period. As we can see, truncation means that the data available is not representative of the general population of the study and therefore introduces bias into the analysis. When this happens, the analysis needs to be adjusted in order to have a more accurate estimate for the population of interest.

4.2.2 Kaplan-Meier estimator

Kaplan-Meier estimator is a non-parametric statistic to estimate the survival rate. It is a piece-wise constant function that measures the rate of survival at a point in time conditional on that the subject has survived until the said time point. The total time span of the study is subdivided into K intervals such that $0 = s_0 < s_1 < \ldots < s_K = \underset{i}{max}(t_i)$. Interval k is defined as $(s_{k-1}, s_k]$. The survival rate for each interval is obtained empirically by calculating the proportion of the number of subjects that remained in the study by the end of the interval, and the total number of subjects at the beginning of the interval. In interval k, it is $\frac{n_k - d_k}{n_k}$, where n_k is the number of subjects in the study at time s_{k-1}, and d_k is the number of subjects that experienced the event during interval k. Since the Kaplan-Meier estimator is the conditional probability for each time period, the survival rate from the current interval needs to be multiplied by the rates from all

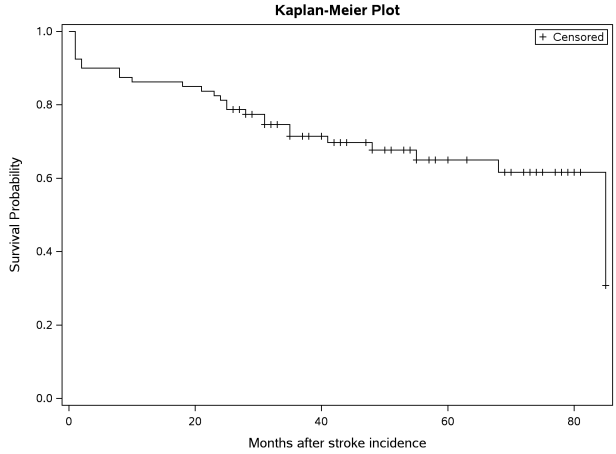

Figure 14: Kaplan-Meier estimator, step-wise function with sharp drop at each time points s_k.

previous intervals to take into account the conditionality. This can be expressed as:

$$\hat{S}(t) = \prod_{t_k < t} (1 - \frac{d_k}{n_k}), \quad k = 1 \dots K$$

This estimator gives a step-wise function with sharp drop at each observed time t_k (as illustrated in figure 14). The finer the intervals, the smoother the curve will become. It is nonetheless not very realistic, and more importantly does not take into account covariates at the individual level.

4.2.3 Cox proportional hazards model

The Cox PH model is a semi-parametric model that describes the hazard function with the covariates. It is the product of two components: a baseline hazard function and the exponential of a linear predictor and can be expressed as

$$h(t) = h_0(t) \exp(\beta^T x) \tag{4.2}$$

where $h_0(t)$ is the baseline hazard function, x is the set of covariates and β is a vector of coefficients for the covariates (Cox, 1972).

One of the most common methods for the baseline hazard function is a non-parametric, piece-wise constant function (Breslow, 1972). It estimates the hazard of the population as a whole, assuming all covariates to be zero. This means that the value of $h_0(t)$ is independent of the individuals. The baseline hazard can be derived from equation (4.1) and a survival function. The survival time span is partitioned in the same way as in the Kaplan-Meier estimator in section 4.2.2. Let the piece-wise constant function be

$$h_0(t) = \lambda_k, \quad \text{for } t \in (s_{k-1}, s_k],\ k = 1, \dots, K. \tag{4.3}$$

As we can see from equation (4.2), in addition to $h_0(t)$, we also have a predictor as a linear combination of the covariates at the individual level that contributes to the total hazard of each individual. This can be extended as the general form of an additive model in equation (2.6) to include terms with fixed effect, random effect, and spatial effect. The predictor is a function that may increase and/or decrease over time. This allows the hazard to also increase and decrease over time. Unlike the Kaplan-Meier estimator which is strictly decreasing, this is more flexible in reflecting the nature of the data.

Consider two subjects i and j, and their respective hazards as stated in equation (4.2). The hazard ratio is

$$\begin{aligned} HR = \frac{h_i(t)}{h_j(t)} &= \frac{h_0(t)\exp(\beta^T x_i)}{h_0(t)\exp(\beta^T x_j)} \\ &= \exp(\beta^T(x_i - x_j)). \end{aligned}$$

The baseline function is the portion that depends on time and is independent of individuals. The hazard ratio is therefore invariant of time. In other words, it remains as the same proportion across time. For this reason, the model is called a proportional hazard model. This is a fundamental assumption of the Cox PH model, and data that is in violation of it should consider an alternative survival model.

4.3 Bayesian survival analysis

4.3.1 Survival model using Bayesian inference

Survival analysis performed within the Bayesian framework gives us posterior results in the form of probability distributions. This gives us more insight into the parameters rather than relying solely on point estimates. However, since the models tend to be more complex, traditional MCMC methods could be computationally expensive and require checking of convergence. INLA as we saw in section 2.2.1 is a much more efficient method.

Since the exponential function component in equation 4.2 is additive, it can easily be incorporated in the framework of latent Gaussian models. In addition, it can then be combined with other model components reflecting the data structure, such as random effects and spatial effects (Moraga, 2019). We will now examine how a Cox PH model can fit into the INLA framework.

4.3.2 Implementation in INLA

As seen in expression 4.3, the baseline hazard function is piece-wise constant. Therefore the hazard function in equation (4.2) during the interval k can be written as

$$\begin{aligned} h(t) &= \lambda_k \exp(\beta^T x) \\ &= \exp(b_k + \beta^T x) = \exp(\eta_k) \quad \text{for } t \in (s_{k-1}, s_k], \quad k = 1, \ldots, K \end{aligned}$$

where $b_k = \ln(\lambda_k)$. So we have $\eta_k = b_k + \beta^T x$. By assigning Gaussian priors to the parameters b_k as well as the vector β, we end up with a Gaussian distributed linear predictor η_k given hyperparameters.

Consider the log-likelihood for the data pair (t, δ) (see section 4.2.1) in interval k, we have

$$\begin{aligned} l = \ln(h(t)S(t)) &= \delta \ln(h(t)) - \int_0^t h(u)du \\ &= \delta\eta_k - (t - s_k)\exp(\eta_k) - \sum_{j=1}^{k-1}(s_{j+1} - s_j)\exp(\eta_j). \end{aligned} \tag{4.4}$$

Notice that this depends on the Gaussian latent field through the vector $\eta_1, \ldots, \eta_k$ and does not fit in the INLA framework. In order for us to use INLA for the Cox PH model, we need to rewrite the dependent variable. Holford (1980) and Laird and Olivier (1981) independently noted that the log-likelihood above is equivalent to the log-likelihood of a Poisson distribution. Equation (4.4) can therefore be regarded as the log-likelihood of k Poisson distributed data points. $k - 1$ of these have the mean $\lambda = (s_{j+1} - s_j)\exp(\eta_j)$ and the observation equals 0. One data point has the mean $\lambda = (t - s_k)\exp(\eta_k)$ with an observation equal to δ (Martino et al., 2011). In other words, if the observation (represented by (t, δ)) is censored, then the Poisson data point would have a value of 0, but if the event is observed, then it would have a value of 1.

To see this, let's consider the log-likelihood of a Poisson distribution $x_i \sim \text{Poisson}(\lambda_i)$

$$l(\lambda; x_1, x_2, \ldots, x_k) = \sum_{i=1}^{k}(-\lambda + x_i \ln(\lambda) - \ln(x_i!)) \tag{4.5}$$

In the case where the there is censoring, we have $x_i = 0, \forall\, i = 1, ..., k$. Hence with the aforementioned respective λs, equation (4.5) becomes

$$l = -(t - s_k)\exp(\eta_k) - \sum_{j=1}^{k-1}(s_{j+1} - s_j)\exp(\eta_j) \tag{4.6}$$

In the case where the event is observed, then we would have one observation with the value of 1 while the others remain 0. Again with the aforementioned respective λs, equation (4.5) becomes

$$\begin{aligned} l &= -(t - s_k)\exp(\eta_k) + \ln((t - s_k)\exp(\eta_k)) - \sum_{j=1}^{k-1}(s_{j+1} - s_j)\exp(\eta_j) \\ &= -(t - s_k)\exp(\eta_k) + \ln(t - s_k) + \eta_k - \sum_{j=1}^{k-1}(s_{j+1} - s_j)\exp(\eta_j) \end{aligned} \tag{4.7}$$

Note that $\ln(t - s_k)$ is a constant and can be disregarded. From equations (4.6) and (4.7), we can see that equation (4.4) is equivalent to (4.5), where the term η_k in (4.7) is controlled by the value δ in the representation in (4.4). Extending this finding to our data pairs. Martino et al. (2011) proposed to rewrite the pairs into K data points that are Poisson distributed which would then allow us to use INLA. These data points have the value 0, except for the position that represents the time interval for t_i which would have a value of 1 if the event was observed.

Now that we see that the Cox PH can fit into the INLA framework. This has the advantage of the flexibility to use different components in the model. For example, random effects and spatial effects can be used just as we do in a regression model. This allows us to work with complex data structures without much additional effort or time.

4.4 Survival analysis results

We are interested in performing survival analysis on the stroke patients. The focus is to find out if there are any geographical variations in the survival rate. Moreover, we would also like to examine other attributes such as age, gender and socioeconomic status of the patients.

Since stroke patients often have other medical conditions, many die of reasons other than stroke. These patients can be quite different from those that died of stroke and we would like to investigate them separately. Additionally, men and women are considered separately since we are interested in seeing if there are specific differences between the genders. Therefore we have 4 sub-populations: women with death by stroke, women with death by other reasons, men with death by stroke, and men with death by other reasons.

In this section, we will first identify what type of censoring and truncation that applies to our data. Next Kaplan-Meier estimator will be presented to give us a peek into the survival rate of the sub-populations. Finally we will use INLA to build Cox proportional hazard models to investigate the baseline hazard as well as the effect of the covariates.

4.4.1 Censoring and truncation

Section 4.2.1 introduced the concept of censoring and truncation. Before doing the analysis, we first examine the data closely so the models can be set up and adjusted appropriately.

Since the data is of patients from a predefined period of 2014-2018 with a follow-up time until the end of 2020, there is censoring in the data. All patients would have been alive at the start of the study period, so there is no left censoring. However, patients who survived beyond 2020 would be right censored.

Additionally, we are interested in cause-specific models, those that died of one cause within the study period are also censored in the model for the other cause since that is the point in time they exit the study. This censoring between the sub-models makes them competing risk models (Hoel, 1972). A competing risk model incorporates more than one event in the model. These events are mutually exclusive and the subjects can only reach at most one of the events, hence the name competing. Once the event occurs, it is terminal and the risk of all other events become zero. Here we have the patients as the subjects and stroke death and other death as the two competing events.

There could also be left truncation. Since the patient population is from the Norwegian stroke registry, only patients that were sent to the hospitals after the onset of the incidence are available. This means that those with extremely short survival time (for example, died at scene of stroke without having the opportunity to reach the healthcare system) are excluded from the study. However, this is very poorly documented and we are unable to estimate the frequency of the scenario. It is natural to assume that this is a very rare event and we will proceed with the analysis without adjustment for truncation.

4.4.2 Kaplan-Meier estimator

We run the Kaplan-Meier method with the stroke patient population. The model (figure 15 (left)) shows a piece-wise constant curve for the survival rate for each time interval given that the patient has survived up until that point. We see that the curve starts at 100% survival rate at time 0 since all patients are alive at the initial point. The interval of the piece-wise function is one month. In month 1 we observe the steepest drop in survival rate, from 100% to about 90%. Since the patient population is from 2014-2018 and we have survival information until the end of 2020, the first censored observation appears around month 24. This is shown on the plot with a + sign. If we follow the curve to the endpoint, we see that the survival rate at month 86 is roughly 50%.

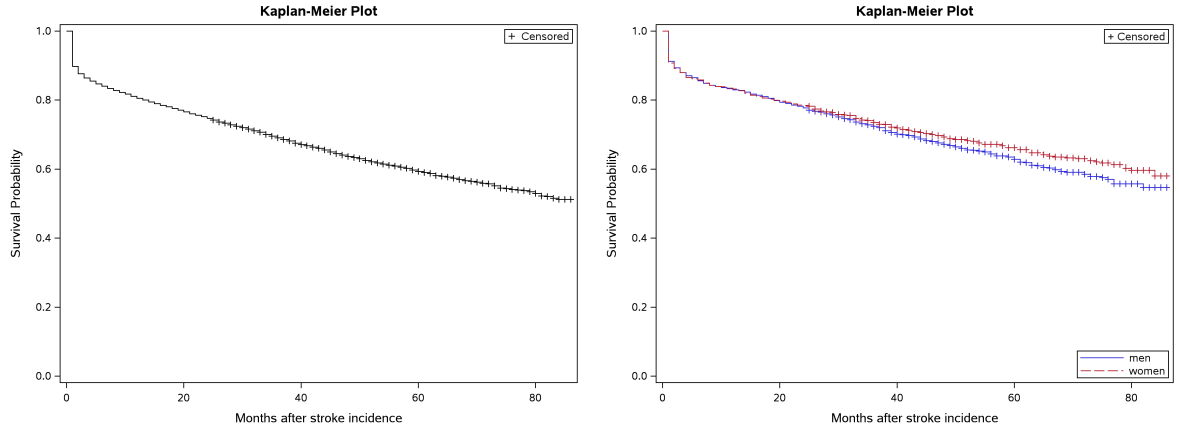

Figure 15: Kaplan-Meier estimator for all stroke patients on the left, and with a stratified sample for men and women separately on the right.

In section 4.1, we created a stratified sample in order to look at the mortality rate by gender without the effect of age. Here we use the same sample for the Kaplan-Meier estimator by gender. Figure 15 (right) shows that male and female stroke patients have very similar survival rates up to 24 months after the onset of stroke. Beyond that, the men's survival rate drops at a faster rate than that of women's.

Since we are interested in seeing the survival rate by the cause of death, we will further split the Kaplan-Meier estimator by those that died of stroke, and those that died of other causes. Figure 16 (left) shows that the sharp drop during the initial months is mostly attributed to stroke death. Also we observe that there are minimal differences in the survival curves for men and women. On the other hand, the survival rate from other causes in figure 16 (right) shows something quite different. There is not an initial drop, but rather a steady decline. Moreover, men have a lower survival rate than women when considering death by other causes than stroke. This explains the difference between genders that we saw earlier in figure 15. Another thing to note in these figures is that the censoring starts much earlier than month 24 as it did in the two previous estimators. This is because when we consider two mutually exclusive events, those who died of one cause are censored at the time of death in the model for the other cause. This event can happen at any point and need not be earliest at the end of the study period (end of 2020). As a result, the censoring naturally constructs the two models that we will explore later - stroke death and other death - into competing risk models.

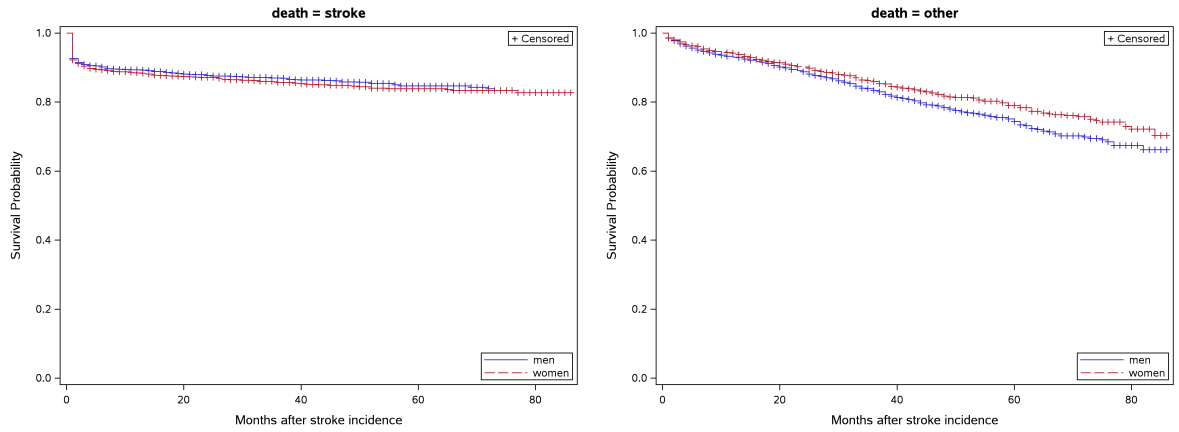

Figure 16: Kaplan-Meier estimator for stroke specific death (left), and other death (right).

4.4.3 Cox proportional hazards model

As an extension of the Kaplan-Meier model, we will now use the Cox PH survival model. In addition to the piece-wise constant function to capture the baseline hazard, the Cox PH model also uses an exponential function with covariates for each individual patient to see which of the attributes have an effect on surviving stroke. The Cox PH model consists of the following covariates (see section 4.1 for more details): age groups, type of stroke [11], time spent in hospital, income[11], education[11], travel time to hospital[11], and place of residence. Categorical variables are set as factors in R, which by default uses the lowest category as the reference category in the model and hence not part of the covariates. For example, income level 1 is the reference category for the income variable, thus only income levels 2-4 are part of the covariates. Moreover, time spent in the hospital (in days) was presented in section 4.1 as categories for the ease of understanding and interpreting the data. In the modelling process, we transform it with the log function in order to minimize the skewness of the data.

The linear predictor η_k is an additive model as seen in equation 2.6 and can easily incorporate different components in the model in the INLA framework. In particular, we model time spent in hospital with RW2 since we know that there is an initial peak and that it is a non-linear relationship. We also use a BYM2 model for the place of residence to observe for spatial effects. A sensitivity analysis was conducted (see appendix A.3) and the values $U = 1$ and $\alpha = 0.01$ were chosen for the precision parameter for PC prior.

For the K baseline hazard function b_k, $k = 1, \ldots, K$, since it is Poisson distributed piece-wise log-constant, we stay with the default setting in INLA and model it with RW1. Furthermore, a PC prior with $U = 0.3$

[11] Categorical variable

and $\alpha = 0.01$ is used for the precision hyperparameter which allows us to control the smoothness of the curve.

The Cox PH model can be written as

$$h(t) = h_0(t) \exp(\beta_0 + \beta_{age}x_{age} + \beta_{type}x_{type} + \beta_{trav}x_{trav} + \beta_{edu}x_{edu} + \beta_{inc}x_{inc} + u_{time} + v_{place})$$

The results for each of the 4 models are presented below first by looking at stroke death for each gender separate, then for other death for each gender.

Stroke death

Women Figure 17 is for female stroke patients who died of stroke. The estimated baseline hazard function on the top left of the figure shows the common hazard rate of dying from stroke for all female patients. Here we use the log transform as it is on a scale that makes it easier to see the details from 5 months onward. We see that the hazard is extremely high immediately after the onset of the stroke but declines steeply and levels out after 5 months.

The rest of the 3 sub-figures are for the exponential function component. The figure on the top right shows the coefficients for the fixed effects. This is also presented in table 4. Age has a positive effect on the hazard; the older the patient, the more likely to die of stroke. When it comes to the type of stroke, since the model uses hemorrhagic stroke as the reference group, a negative coefficient for the ischemic stroke type means that it is less hazardous compared to hemorrhagic stroke. Both of the above-mentioned variables have credible intervals that do not cross 0, which means that they are significant hazards. The next 4 variables are for the different lengths of travel time to the local hospital (dist2 - dist5) and all of them have credible interval that cross 0. Hence it is not a significant hazard to stroke death. For both education and income levels, we see a slight decreasing trend, which suggests that the higher the education and income level, the lower the hazard. However, for the education level, only the last category - university level and above - is significant due to the strictly negative credible interval. This means that patients with university or above education have lower hazards than those with the reference group: obligatory education level. For the income level, all 3 categories, which collectively refer to annual household income after tax above 385 000 NOK, are negative and significant, but the credible intervals for these 3 overlap quite a lot, so we can not say that they have different effects from each other. However, since the CIs do not cross 0, we can say that patients with income higher than 385 000 NOK have decreased hazard compared to those with income lower than the said amount.

In section 4.1, we saw that the number of days spent in the hospital does not have a simple linear relationship with the mortality rate. Therefore we have chosen to model it as a non-linear random effect with

a random walk 2 model. The figure on the lower left shows that similarly to what we saw in section 4.1 there is a non-linear pattern with a decrease at the beginning, and then an increase after a certain point. Additionally, as part of the model exploration and validation, another model with time spent in hospital as a linear effect was also done to be compared with the results of the non-linear effect. DIC is 20206 with a linear effect as opposed to 20097 when assuming a non-linear effect. This confirms that the effect is better captured as a non-linear function.

Lastly, we investigate the geographical effect in the survival model. As in the spatial model, we use a BYM2 model (see section 3.3.4) to capture both the unstructured and the structured spatial effect of the RAs for the patients. The two lower right figures show that the majority of the CIs clearly cover 0 hence there are no spatial effects. The only exception is that RA12 (Østfold) shows sign of increased hazard. As shown in appendix A.3, the result can depend on the prior chosen. The specific prior that we used gives us a positive effect for one RA, but it is nonetheless still very close to 0 and in fact is not significant with other U values for the prior. We therefore take note of it but conclude that there are no geographical effects to the hazard of stroke death for female stroke patients.

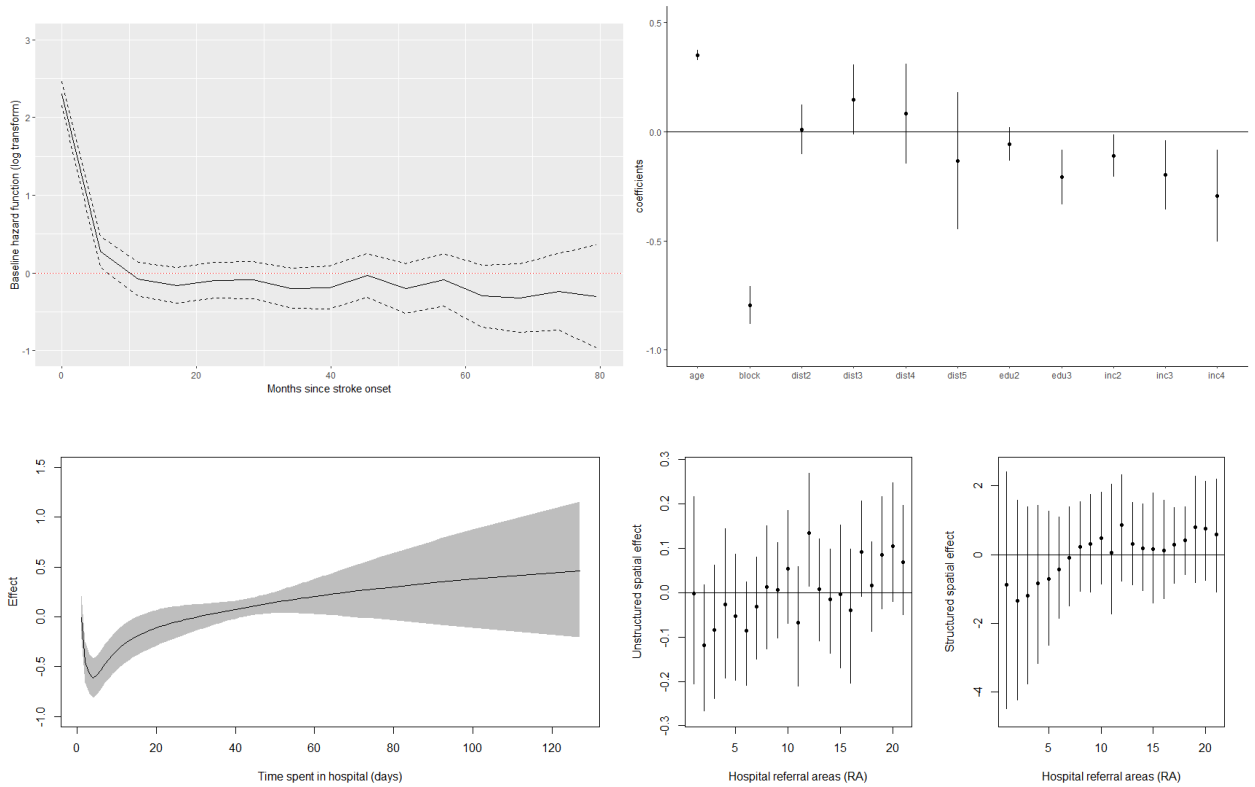

Figure 17: Death caused by stroke for women.
(Top left) The baseline hazard function in the log scale. (Top right) Coefficients of the fixed effects: age groups, blockage (ischemic stroke), distance (travel time to hospital in categories), education, and income. (Bottom left) Random effect of time spent in hospital. (Bottom right) Spatial effects from the BYM2 model component.
The 2 line graphs on the left have solid line showing the posterior mean and the dotted lines or shaded area for 95% CI. The 3 plots on the right have the black circle showing the posterior mean and the line segment for the 95% CI.

Men For male stroke patients to survive death by stroke, the baseline hazard function exhibits a similar pattern as their female counterparts. Moreover, a combined model with gender as a fixed effect also shows that gender does not have an effect on hazard rate. If we look at the other plots for the covariates, we see that most of them have very similar results as in the model for women. The only notable difference is income. In the case of men, it is more pronounced that the higher the income, the lower the hazard since CIs do not overlap as much as in the model for female patients.

Women, stroke death		
Covariates	Mean	95%CI
(Intercept)	-5.88	(-6.19, -5.57)
Age group	0.35	(0.33, 0.37)
Block (ischemic)	-0.80	(-0.88, -0.71)
Travel time (31-60 min)	0.01	(-0.10, 0.12)
Travel time (61-90 min)	0.15	(-0.01, 0.31)
Travel time (91-120 min)	0.08	(-0.15, 0.31)
Travel time (120+ min)	-0.13	(-0.45, 0.18)
Education (high school)	-0.06	(-0.14, 0.02)
Education (university)	-0.21	(-0.34, -0.08)
Income (385-570K)	-0.11	(-0.21, -0.01)
Income (570-795K)	-0.20	(-0.36, -0.04)
Income (795K+)	-0.30	(-0.51, -0.08)

Men, stroke death		
Covariates	Mean	95%CI
(Intercept)	-5.71	(-6.04, -5.38)
Age group	0.34	(0.32, 0.36)
Block (ischemic)	-0.83	(-0.92, -0.73)
Travel time (31-60 min)	0.02	(-0.11, 0.14)
Travel time (61-90 min)	-0.08	(-0.27, 0.11)
Travel time (91-120 min)	-0.09	(-0.36, 0.19)
Travel time (120+ min)	-0.17	(-0.49, 0.15)
Education (high school)	-0.04	(-0.13, 0.06)
Education (university)	-0.20	(-0.33, -0.07)
Income (385-570K)	-0.19	(-0.28, -0.09)
Income (570-795K)	-0.39	(-0.53, -0.25)
Income (795K+)	-0.61	(-0.80, -0.42)

Table 4: Posterior mean and 95% credible interval for the fixed effects in the survival models for stroke death.

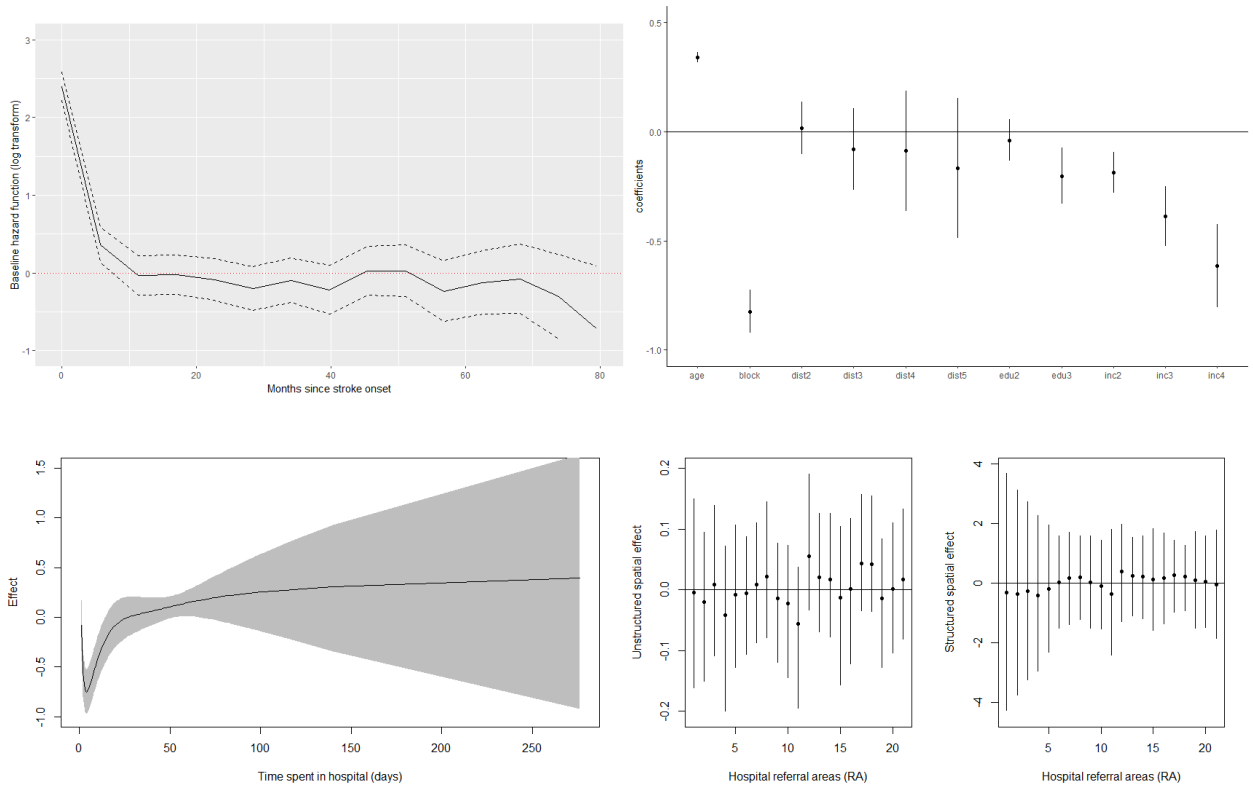

Figure 18: Death caused by stroke for men. See figure 17 for detailed description.

Other death

Women In a similar fashion, we can interpret the output shown in figure 19 of the model for female stroke patients who died of reasons other than stroke. The baseline hazard function does not have the initial hike, and has an even and stable hazard throughout the 80 months of the study period. This means that the majority of the hazard is described by the covariates in the exponential model. Just like in the model for stroke death, higher age also suggests heightened hazard. Ischemic stroke patients have a lower hazard than hemorrhagic stroke patients, even though this concerns those who died of reasons other than stroke. Travel time to hospital still has no significant effect. Medium and high level education have lower hazards compared to those with obligatory level of education. Those with income levels 385 000 - 795 000 NOK have lower hazards compared to those with income less than 385 000 NOK. There are a couple of RAs with CI above or below 0, but are all very close. If we see the spatial effect as a whole, it is safe to say that there are no geographical effects to hazard.

We see that the major difference between the models for stroke death and other death for women is the underlying baseline hazard. There is a very high hazard for stroke death during the initial phase of the stroke, but once a patient survives beyond 5 months, the baseline hazard is relatively flat. In comparison, there is not such initial hazard for death by other causes. Most of the covariates have similar effects in both models.

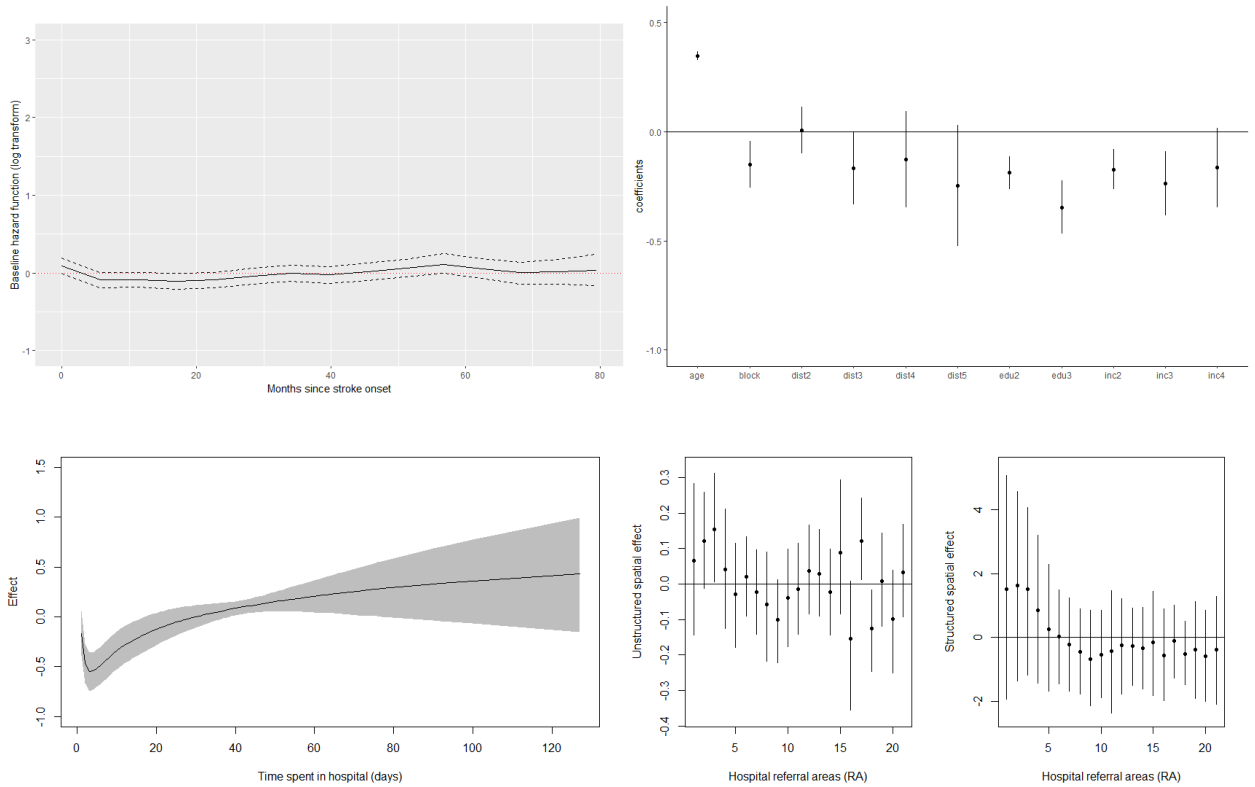

Figure 19: Death caused by reasons other then stroke for women. See figure 17 for detailed description.

Men Lastly, we look at the male stroke patients to survive death caused by reasons other than stroke. The baseline hazard is flat from the onset, as their female counterparts. For the covariates, 60-90 minutes travel time to a local hospital has lower hazard than the others, but the credible interval is very close to 0. Since none of the other 4 travel time categories are significant, even though this is technically significant, the model most likely would not lose much strength without it. For the socioeconomic measures, higher education and higher income have lower hazard.

When comparing male patients for stroke death and other death, we see the same pattern as for female patients that the biggest difference is in the baseline hazard, and the covariates have very similar effects.

Women, other death		
Covariates	Mean	95%CI
(Intercept)	-5.23	(-5.50, -4.96)
Age group	0.35	(0.33, 0.37)
Block (ischemic)	-0.15	(-0.26, -0.05)
Travel time (31-60 min)	0.01	(-0.10, 0.11)
Travel time (61-90 min)	-0.17	(-0.34, 0.00)
Travel time (91-120 min)	-0.13	(-0.35, 0.09)
Travel time (120+ min)	-0.25	(-0.53, 0.03)
Education (high school)	-0.19	(-0.26, -0.11)
Education (university)	-0.35	(-0.47, -0.22)
Income (385-570K)	-0.17	(-0.26, -0.08)
Income (570-795K)	-0.24	(-0.39, -0.09)
Income (795K+)	-0.16	(-0.35, 0.02)

Men, other death		
Covariates	Mean	95%CI
(Intercept)	-5.06	(-5.32, -4.81)
Age group	0.36	(0.34, 0.38)
Block (ischemic)	-0.26	(-0.36, -0.15)
Travel time (31-60 min)	0.01	(-0.09, 0.11)
Travel time (61-90 min)	-0.18	(-0.33, -0.02)
Travel time (91-120 min)	0.07	(-0.13, 0.28)
Travel time (120+ min)	0.08	(-0.15, 0.31)
Education (high school)	-0.10	(-0.18, -0.03)
Education (university)	-0.25	(-0.36, -0.15)
Income (385-570K)	-0.29	(-0.36, -0.21)
Income (570-795K)	-0.45	(-0.56, -0.34)
Income (795K+)	-0.52	(-0.66, -0.37)

Table 5: Posterior mean and 95% credible interval for the fixed effects in the survival models for other death.

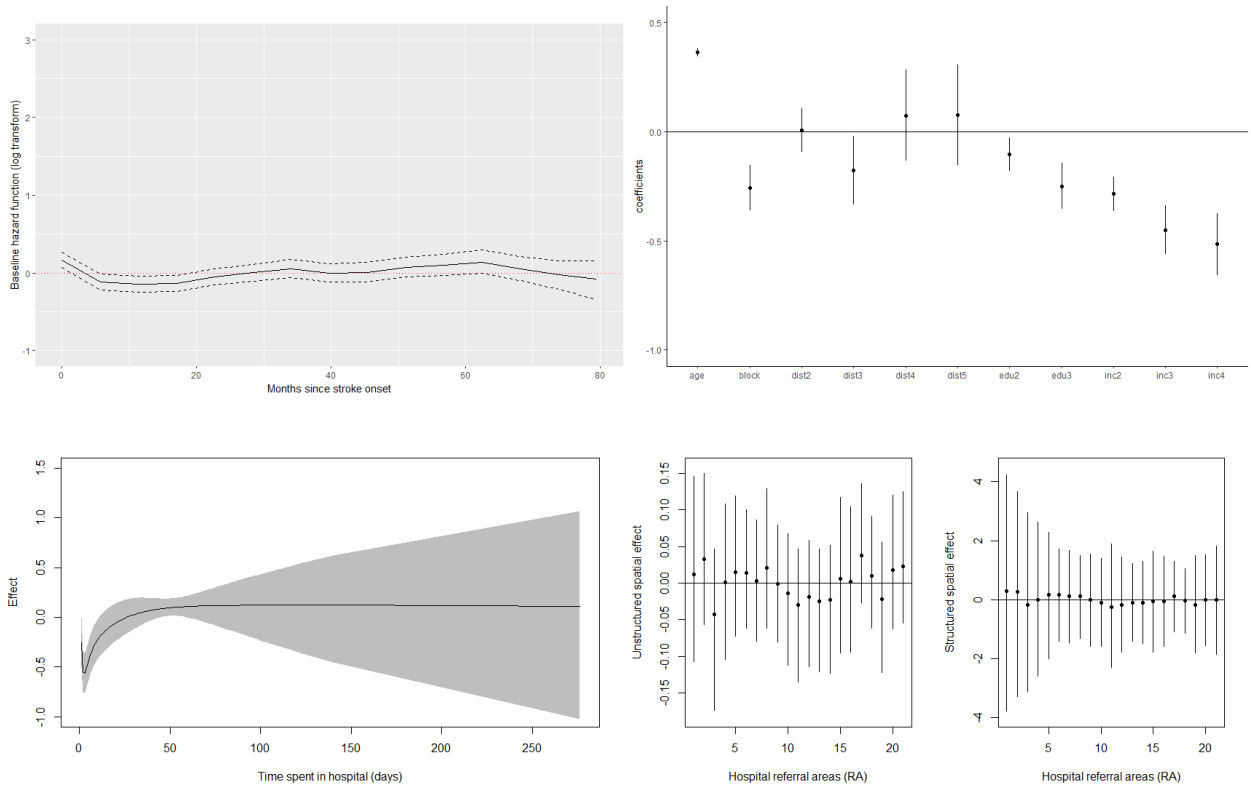

Figure 20: eath caused by reasons other then stroke for men. See figure 17 for detailed description.

4.5 Remarks on risk factors in stroke mortality rates

The most notable result from the models is that geographical differences do not affect the survival of stroke patients. This confirms that stroke patients receive equal quality healthcare services regardless of where they live in Norway. In addition, we also learn from the models that there is a heightened baseline hazard of stroke death within the first 5 months of the stroke incidence. This is intuitive and to be expected as stroke is an acute condition that is often quite fatal.

Aside from the baseline hazard that applies universally to all patients, each patient has specific attributes that may increase or decrease their hazard rate. If we consider the effect of patient demographics - age and gender - older patients have higher hazard while gender does not seem to play a major role.

When it comes to socioeconomic status, patients with an education level at or above university have a lower hazard of stroke death compared to those with an obligatory educational level (grade 10). In addition, patients with high school education also have a lower hazard compared to those with obligatory education when it comes to other death. This is true for both men and women. The effect of income level is slightly different for men and women. Female patients with household income greater than 385 000 NOK have a lower hazard, for both stroke death and other death, than those with lower income. This is also true for male patients, and in addition those with income greater than 570 000 NOK have an even lower hazard than those with income less than 570 000 NOK. Such socioeconomic differences in stroke mortality is also observed in other international studies (Avan et al., 2019; Addo et al., 2012).

In summary, gender and geographical placement have no effect on the hazard rate while type of stroke, education and income do. Furthermore, the hazard of stroke death is higher during the first 5 months of the incidence while the hazard of other death remains relatively unchanged over time.

5 Conclusion

In the Norwegian healthcare system, one of the mandates is to provide equal and high quality healthcare services to the residents no matter where they live and what their social status may be. In this thesis, we looked at stroke patients and tackled two main questions: are there any geographical differences in the stroke incidence rate in Norway, and are place of residence and socioeconomic status risk factors for stroke mortality?

We used two different methods to examine the stroke incidence rate. Hypothesis tests showed that many geographical areas have significantly different stroke incidence rates. However, when we examined the data with a Bayesian spatial model, we saw that places of residence have little to no effect on incidence rates. The difference in the result is due to the analysis methods and the underlying assumptions. We see the advantage of using a more sophisticated modelling technique that requires fewer assumptions and provides a more data-driven result. This highlights the importance of careful consideration of the analytical methods employed and whether they are suitable given the data and the situation.

For the mortality rate of stroke patients, we used the Cox PH model within the Bayesian framework to explore the effect of geographical differences and socioeconomic status. This allowed us to measure if there are significant variations in the health care provided. We see that place of residence has no effect on the mortality rate, while higher income and education may reduce mortality.

Stroke mortality is a complex matter and may include many interrelated risk factors. Here we have only explored some basic metrics, but there are many others that can be included. For example, numerous studies have shown that comorbidity (such as, but not limited to, high blood pressure and diabetes) and stroke severity are also significant factors when it comes to stroke mortality (Rost et al., 2016; Phan et al., 2016; Sarker et al., 2008). Moreover, as stroke is a serious acute condition, the timeliness and response from the medical team can also play an important role. This in particular would enrich our understanding of the geographical variation in the healthcare system. The advantage of the Bayesian Cox PH model is that we can incorporate other factors as either fixed effects or random effects to capture extra layers that may be present in the data. This can be done with minimal effort and can provide much more insight into the study subject. Additionally with the INLA method, the efficiency of the model performance facilitates exploration of the intricate relationship between the risk factors. Further study with an extended list of potential risk factors could uncover more details or explain the results we have seen from the analysis in this thesis.

A Appendix

A.1 Regional health authorities (RHA) and hospital referral areas (RA)

Number	RHA	RA
1	North	Finnmark
2	North	Tromsø
3	North	Nordland
4	North	Helgeland
5	Central	Nord-Trøndelag
6	Central	St. Olaves
7	Central	Møre og Romsdal
8	West	Førde
9	West	Bergen
10	West	Fonna
11	West	Stavanger
12	South-east	Østfold
13	South-east	Akershus
14	South-east	OUS (Oslo University Hospital)
15	South-east	Lovisenberg
16	South-east	Diakonhjem
17	South-east	Innlandet
18	South-east	Vest Viken
19	South-east	Vestfold
20	South-east	Telemark
21	South-east	Sørlandet

A.2 Neighbourhood map

In order to create a spatial model, we need to capture the information that is normally presented on a map into a data format. Each area on the map is assigned a number, and we note down the number of neighbouring areas, as well as the corresponding area numbers for the neighbours. The neighbourhood map for Norwegian hospital referral areas is as follows. First line indicates total number of areas, in this case 21. As an example, the way to read the next line is: "RA 1 (Finnmark) has 1 neighbour which is RA 2 (Tromsø)".

```
21
 1    1   2
 2    2   1  3
 3    2   2  4
 4    2   3  5
 5    2   4  6
 6    3   5  7  17
 7    3   6  17
 8    4   7  9  17 18
 9    3   8  10  18
10    5   9  11  18  20  21
11    2   10  21
12    3   13  18  19
13    6   12  14  15  16  17 18
14    5   13  15  16  17  18
15    3   13  14  16
16    4   13  14  15  18
17    6   6  7  8  13  14  18
18   10   8  9  10  12  13  14  16  17  19  20
19    3   12  18  20
20    4   10  18  19  21
21    3   10  11  20
```

A.3 Sensitivity analysis for PC prior

This table shows the result of the sensitivity analysis with different U values in the PC prior for the precision parameter in the BYM2 model, while keeping α fixed at 0.01. The first 21 rows represent the unstructured spatial random effect for the hospital referral areas and last 21 rows represent the structured spatial random effect for the RAs. The numbers in the table represent the posterior mean of the effect with the 95% credible interval in the parenthesis. Notice that almost all credible intervals cover zero except for the unstructured effect for RA 12 with $U > 0.5$. So the analysis is robust to the prior choices.

RA	U=0.05	U=0.1	U=0.5	U=1	U=2	U=5
Unstructured spatial effect						
1	-0.03 (-0.17, 0.10)	-0.02 (-0.19, 0.15)	-0.00 (-0.21, 0.21)	-0.00 (-0.21, 0.22)	0.00 (-0.21, 0.22)	0.00 (-0.21, 0.22)
2	-0.06 (-0.19, 0.02)	-0.10 (-0.23, 0.02)	-0.12 (-0.26, 0.02)	-0.12 (-0.27, 0.02)	-0.12 (-0.27, 0.02)	-0.12 (-0.27, 0.02)
3	-0.05 (-0.17, 0.04)	-0.07 (-0.21, 0.05)	-0.08 (-0.24, 0.06)	-0.08 (-0.24, 0.06)	-0.08 (-0.24, 0.06)	-0.09 (-0.24, 0.06)
4	-0.02 (-0.14, 0.07)	-0.03 (-0.17, 0.11)	-0.03 (-0.19, 0.14)	-0.03 (-0.19, 0.14)	-0.03 (-0.19, 0.15)	-0.03 (-0.20, 0.15)
5	-0.03 (-0.13, 0.05)	-0.04 (-0.17, 0.07)	-0.05 (-0.19, 0.08)	-0.05 (-0.20, 0.09)	-0.05 (-0.20, 0.09)	-0.05 (-0.20, 0.09)
6	-0.04 (-0.14, 0.03)	-0.06 (-0.18, 0.03)	-0.08 (-0.20, 0.02)	-0.09 (-0.21, 0.02)	-0.09 (-0.21, 0.02)	-0.09 (-0.21, 0.02)
7	-0.01 (-0.10, 0.06)	-0.02 (-0.13, 0.07)	-0.03 (-0.15, 0.08)	-0.03 (-0.15, 0.08)	-0.03 (-0.15, 0.08)	-0.03 (-0.15, 0.08)
8	0.01 (-0.07, 0.09)	0.01 (-0.10, 0.12)	0.01 (-0.12, 0.15)	0.01 (-0.13, 0.15)	0.01 (-0.13, 0.15)	0.01 (-0.13, 0.15)
9	0.01 (-0.06, 0.08)	0.01 (-0.09, 0.10)	0.01 (-0.10, 0.11)	0.01 (-0.10, 0.11)	0.01 (-0.10, 0.11)	0.01 (-0.10, 0.12)
10	0.02 (-0.04, 0.12)	0.04 (-0.06, 0.15)	0.05 (-0.07, 0.18)	0.05 (-0.07, 0.19)	0.05 (-0.07, 0.19)	0.05 (-0.07, 0.19)
11	-0.02 (-0.13, 0.06)	-0.04 (-0.17, 0.06)	-0.06 (-0.21, 0.06)	-0.07 (-0.21, 0.06)	-0.07 (-0.21, 0.06)	-0.07 (-0.22, 0.06)
12	0.06 (-0.01, 0.18)	0.10 (-0.00, 0.23)	**0.13 (0.01, 0.26)**	**0.13 (0.01, 0.27)**	**0.14 (0.02, 0.27)**	**0.14 (0.02, 0.27)**
13	0.01 (-0.06, 0.08)	0.01 (-0.09, 0.11)	0.01 (-0.11, 0.12)	0.01 (-0.11, 0.12)	0.01 (-0.11, 0.12)	0.01 (-0.11, 0.12)
14	-0.00 (-0.08, 0.07)	-0.01 (-0.11, 0.09)	-0.01 (-0.13, 0.10)	-0.01 (-0.14, 0.10)	-0.02 (-0.14, 0.10)	-0.02 (-0.14, 0.10)
15	0.01 (-0.09, 0.10)	0.00 (-0.13, 0.13)	-0.00 (-0.16, 0.15)	-0.00 (-0.17, 0.15)	-0.00 (-0.17, 0.15)	-0.01 (-0.17, 0.16)
16	-0.01 (-0.11, 0.07)	-0.02 (-0.15, 0.09)	-0.04 (-0.20, 0.10)	-0.04 (-0.20, 0.10)	-0.04 (-0.21, 0.10)	-0.04 (-0.21, 0.10)
17	0.04 (-0.02, 0.14)	0.07 (-0.02, 0.18)	0.09 (-0.01, 0.20)	0.09 (-0.01, 0.21)	0.09 (-0.01, 0.21)	0.09 (-0.01, 0.21)
18	0.01 (-0.05, 0.08)	0.01 (-0.07, 0.10)	0.02 (-0.09, 0.11)	0.02 (-0.09, 0.11)	0.02 (-0.09, 0.12)	0.02 (-0.09, 0.12)
19	0.04 (-0.03, 0.14)	0.06 (-0.04, 0.18)	0.08 (-0.04, 0.21)	0.08 (-0.04, 0.22)	0.09 (-0.04, 0.22)	0.09 (-0.04, 0.22)
20	0.04 (-0.02, 0.15)	0.08 (-0.03, 0.20)	0.10 (-0.02, 0.24)	0.10 (-0.02, 0.25)	0.11 (-0.02, 0.25)	0.11 (-0.02, 0.25)
21	0.03 (-0.04, 0.13)	0.05 (-0.05, 0.17)	0.07 (-0.05, 0.19)	0.07 (-0.05, 0.20)	0.07 (-0.05, 0.20)	0.07 (-0.05, 0.20)
Structured spatial effect						
1	-1.39 (-5.41, 2.82)	-1.26 (-5.03, 2.40)	-0.93 (-4.56, 2.42)	-0.88 (-4.50, 2.42)	-0.85 (-4.46, 2.42)	-0.83 (-4.44, 2.43)
2	-1.63 (-5.05, 2.21)	-1.63 (-4.73, 1.62)	-1.38 (-4.29, 1.59)	-1.34 (-4.23, 1.59)	-1.32 (-4.20, 1.59)	-1.30 (-4.18, 1.59)
3	-1.43 (-4.43, 1.92)	-1.44 (-4.17, 1.41)	-1.23 (-3.81, 1.38)	-1.20 (-3.76, 1.39)	-1.18 (-3.74, 1.39)	-1.17 (-3.71, 1.39)
4	-1.06 (-3.68, 1.75)	-1.05 (-3.50, 1.43)	-0.87 (-3.22, 1.43)	-0.84 (-3.18, 1.44)	-0.82 (-3.16, 1.44)	-0.81 (-3.14, 1.44)
5	-0.78 (-2.95, 1.51)	-0.82 (-2.86, 1.29)	-0.72 (-2.68, 1.28)	-0.71 (-2.65, 1.28)	-0.70 (-2.64, 1.28)	-0.69 (-2.62, 1.28)
6	-0.37 (-1.95, 1.26)	-0.45 (-1.94, 1.14)	-0.45 (-1.88, 1.11)	-0.44 (-1.87, 1.10)	-0.44 (-1.86, 1.10)	-0.44 (-1.85, 1.10)
7	-0.01 (-1.56, 1.56)	-0.06 (-1.53, 1.47)	-0.09 (-1.51, 1.40)	-0.10 (-1.50, 1.39)	-0.10 (-1.50, 1.38)	-0.10 (-1.50, 1.38)
8	0.30 (-1.13, 1.71)	0.28 (-1.07, 1.64)	0.23 (-1.08, 1.55)	0.22 (-1.08, 1.53)	0.21 (-1.08, 1.53)	0.21 (-1.08, 1.52)
9	0.40 (-1.20, 1.96)	0.39 (-1.09, 1.88)	0.32 (-1.09, 1.77)	0.31 (-1.09, 1.75)	0.30 (-1.09, 1.74)	0.29 (-1.09, 1.73)
10	0.54 (-1.04, 2.04)	0.56 (-0.88, 1.96)	0.49 (-0.87, 1.84)	0.48 (-0.87, 1.82)	0.48 (-0.87, 1.81)	0.47 (-0.87, 1.80)
11	0.26 (-1.81, 2.36)	0.18 (-1.74, 2.21)	0.07 (-1.74, 2.07)	0.05 (-1.73, 2.05)	0.05 (-1.73, 2.04)	0.04 (-1.73, 2.04)
12	0.84 (-0.99, 2.52)	0.93 (-0.77, 2.48)	0.87 (-0.77, 2.35)	0.86 (-0.77, 2.33)	0.85 (-0.77, 2.31)	0.85 (-0.78, 2.31)
13	0.42 (-0.95, 1.73)	0.40 (-0.86, 1.65)	0.32 (-0.87, 1.53)	0.31 (-0.88, 1.51)	0.30 (-0.88, 1.50)	0.29 (-0.88, 1.50)
14	0.32 (-1.09, 1.71)	0.28 (-1.03, 1.62)	0.19 (-1.05, 1.50)	0.18 (-1.05, 1.48)	0.17 (-1.05, 1.47)	0.17 (-1.06, 1.47)
15	0.32 (-1.41, 2.04)	0.27 (-1.38, 1.94)	0.17 (-1.42, 1.81)	0.15 (-1.42, 1.79)	0.14 (-1.42, 1.78)	0.14 (-1.42, 1.77)
16	0.30 (-1.24, 1.84)	0.24 (-1.22, 1.73)	0.13 (-1.27, 1.60)	0.11 (-1.27, 1.58)	0.10 (-1.28, 1.57)	0.09 (-1.28, 1.56)
17	0.33 (-0.90, 1.52)	0.34 (-0.82, 1.47)	0.30 (-0.83, 1.38)	0.29 (-0.83, 1.37)	0.29 (-0.83, 1.36)	0.29 (-0.83, 1.36)
18	0.49 (-0.71, 1.59)	0.49 (-0.57, 1.52)	0.42 (-0.58, 1.40)	0.41 (-0.58, 1.39)	0.40 (-0.58, 1.38)	0.40 (-0.58, 1.37)
19	0.80 (-1.03, 2.49)	0.88 (-0.81, 2.45)	0.81 (-0.81, 2.31)	0.80 (-0.81, 2.29)	0.79 (-0.81, 2.27)	0.79 (-0.81, 2.26)
20	0.74 (-0.98, 2.32)	0.82 (-0.76, 2.27)	0.76 (-0.75, 2.15)	0.75 (-0.75, 2.13)	0.75 (-0.75, 2.12)	0.74 (-0.75, 2.11)
21	0.62 (-1.31, 2.45)	0.65 (-1.12, 2.37)	0.59 (-1.10, 2.23)	0.58 (-1.10, 2.20)	0.57 (-1.10, 2.19)	0.57 (-1.10, 2.18)

References

J. Addo, L. Ayerbe, K. M. Mohan, S. Crichton, A. Sheldenkar, R. Chen, C. D. Wolfe, and C. McKevitt. Socioeconomic status and stroke. *Stroke*, 43(4):1186–1191, 2012. URL `https://www.ahajournals.org/doi/abs/10.1161/STROKEAHA.111.639732`.

A. Avan, H. Digaleh, M. D. Napoli, S. Stranges, R. Behrouz, G. Shojaeianbabaei, A. Amiri, R. Tabrizi, N. Mokhber, J. D. Spence, and M. R. Azarpazhooh. Socioeconomic status and stroke incidence, prevalence, mortality, and worldwide burden: an ecological analysis from the global burden of disease study 2017. *BMC Medicine*, 17(191), 2019. doi: https://doi.org/10.1186/s12916-019-1397-3.

T. Bayes. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. of the Royal Soc. of London*, 53:370–418, 1763.

E. J. Benjamin, M. J. Blaha, S. E. Chiuve, M. Cushman, S. R. Das, R. Deo, S. D. de Ferranti, J. Floyd, M. Fornage, C. Gillespie, C. R. Isasi, M. C. Jiménez, L. C. Jordan, S. E. Judd, D. Lackland, J. H. Lichtman, L. Lisabeth, S. Liu, C. T. Longenecker, R. H. Mackey, K. Matsushita, D. Mozaffarian, M. E. Mussolino, K. Nasir, R. W. Neumar, L. Palaniappan, D. K. Pandey, R. R. Thiagarajan, M. J. Reeves, M. Ritchey, C. J. Rodriguez, G. A. Roth, W. D. Rosamond, C. Sasson, A. Towfighi, C. W. Tsao, M. B. Turner, S. S. Virani, J. H. Voeks, J. Z. Willey, J. T. Wilkins, J. H. Wu, H. M. Alger, S. S. Wong, and P. Muntner. Heart disease and stroke statistics—2017 update: A report from the American heart association. *Circulation*, 135 (10):e146–e603, 2017. URL `https://www.ahajournals.org/doi/abs/10.1161/CIR.0000000000000485`.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B, Methodological*, 57(1):289–300, 1995. ISSN 0035-9246.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B, Methodological*, 36(2):192–236, 1974. ISSN 0035-9246.

J. Besag, J. York, and A. Molli. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20, 1991. ISSN 0020-3157.

M. Blangiardo and M. Cameletti. *Spatial and spatio-temporal Bayesian models with R-INLA*. Wiley, Chichester, England, 2015. ISBN 1-118-95019-4.

N. E. Breslow. Discussion of the paper by D. R. Cox. *Journal of the Royal Statistical Society. Series B*, 34: 187–220, 1972.

Center for Clinical Documentation and Evaluation (SKDE). Healthcare atlas for the elderly in Norway, 2017. URL https://www.skde.no/helseatlas/en/v2/eldre/.

Centers for Disease Control and Prevention (CDC). Abuot stroke. URL https://www.cdc.gov/stroke/about.htm#:~:text=Download%20Image%5BJPG%5D,term%20disability%2C%20or%20even%20death.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B, Methodological*, 34(2):187–220, 1972. ISSN 0035-9246.

C. Dean, M. Ugarte, and A. Militino. Detecting interaction between random region and fixed age effects in disease mapping. *Biometrics*, 57:197–202, 04 2001.

E. S. Donkor. Stroke in the 21st century: A snapshot of the burden, epidemiology, and quality of life. *Stroke research and treatment*, 3238165, 2018. doi: https://doi.org/10.1155/2018/3238165.

H. Ellekjær, H. Fjærtoft, B. Indredavik, B. Mørch, R. Skogseth-Stephani, and T. Varmdal. Norsk hjerneslagregister – Årsrapport 2015 med plan for forbedringstiltak, 2016. URL https://stolav.no/Medisinskekvalitetsregistre/Norsk-hjerneslagregister/%C3%85rsrapport%20Norsk%20hjerneslagregister%202015.pdf.

S. M. Fernando, D. Qureshi, R. Talarico, P. Tanuseputro, D. Dowlatshahi, M. M. Sood, E. E. Smith, M. D. Hill, V. A. McCredie, D. C. Scales, S. W. English, B. Rochwerg, and K. Kyeremanteng. Intracerebral hemorrhage incidence, mortality, and association with oral anticoagulation use. *Stroke*, 52(5):1673–1681, 2021. doi: 10.1161/STROKEAHA.120.032550. URL https://www.ahajournals.org/doi/abs/10.1161/STROKEAHA.120.032550.

H. Fjærtoft, B. Indredavik, B. Mørch, A. Phan, R. Skogseth-Stephani, and T. Varmdal. Norsk hjerneslagregister – Årsrapport 2016 med plan for forbedringstiltak, 2017. URL https://stolav.no/Medisinskekvalitetsregistre/Norsk-hjerneslagregister/%C3%85rsrapport2016-Norsk-hjerneslagregister.pdf.

H. Fjærtoft, B. Indredavik, B. Mørch, A. Phan, R. Skogseth-Stephani, K. K. Halle, and T. Varmdal. Norsk hjerneslagregister – Årsrapport 2017 med plan for forbedringstiltak, 2018. URL https://stolav.no/Medisinskekvalitetsregistre/Norsk-hjerneslagregister/%C5rsrapport_Norsk_hjerneslagregister%202017.pdf.

H. Fjærtoft, B. Indredavik, B. Mørch, R. Skogseth-Stephani, K. K. Halle, and T. Varmdal. Norsk hjerneslagregister – Årsrapport 2018 med plan for forbedringstiltak, 2019. URL https://stolav.no/Documents/Revidert_%C3%85rsrapport%202018_NHR.pdf.

M. Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675–701, 1937. ISSN 0162-1459.

A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. *Bayesian Data Analysis (3rd ed.)*. Chapman & Hall/CRC., 2013.

V. Gómez-Rubio. *Bayesian Inference with INLA*. Chapman Hall/CRC Press. Boca Raton, FL., 2020.

D. G. Hoel. A representation of mortality data by competing risks. *Biometrics*, 28(2):475–488, 1972. ISSN 0006-341X.

T. R. Holford. The analysis of rates and of survivorship using log-linear models. *Biometrics*, 36(2):299–305, 1980. ISSN 0006-341X.

B. Indredavik, H. Fjærtoft, H. Ellekjær, R. Skogseth-Stephani, T. Varmdal, and B. Mørch. Norsk hjerneslagregister – Årsrapport 2014 med plan for forbedringstiltak, 2015. URL `https://stolav.no/Medisinskekvalitetsregistre/Norsk-hjerneslagregister/%C3%85rsrapport%20Norsk%20hjerneslagregister%202014%20per%2022.10.2015.pdf`.

H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and physical sciences*, 186(1007):453–461, 1946. ISSN 0080-4630.

A. S. Kim, E. Cahill, and N. T. Cheng. Global stroke belt. *Stroke*, 46(12):3564–3570, 2015. URL `https://www.ahajournals.org/doi/abs/10.1161/STROKEAHA.115.008226`.

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22 (1):79–86, 1951. ISSN 0003-4851.

N. Laird and D. Olivier. Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, 76(374):231–240, 1981. ISSN 0162-1459.

B. G. Leroux, X. Lei, and N. E. Breslow. Estimation of disease rates in small areas: A new mixed model for spatial dependence. 2000.

L. Martin, R. Leblanc, and N. K. Toan. Tables for the Friedman rank test. *Canadian Journal of Statistics*, 21(1):39–43, 1993.

S. Martino, R. Akerkar, and H. Rue. Approximate Bayesian inference for survival models. *Scandinavian Journal of Statistics*, 38(3):514–528, 2011. doi: http://www.jstor.org/stable/23015578.

P. Moraga. *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. Chapman Hall/CRC Biostatistics Series, 2019.

N. N. Nyi. Easy way to learn standardization : direct and indirect methods. *The Malaysian Journal of Medical Sciences*, 7(1):10–15, 2000.

T. G. Phan, B. Clissold, J. Ly, H. Ma, C. Moran, and V. Srikanth. Stroke severity and comorbidity index for prediction of mortality after ischemic stroke from the virtual international stroke trials archive–acute collaboration. *Journal of Stroke and Cerebrovascular Diseases*, 25(4):835–842, 2016. ISSN 1052-3057. doi: https://doi.org/10.1016/j.jstrokecerebrovasdis.2015.12.016. URL `https://www.sciencedirect.com/science/article/pii/S1052305715006837`.

A. Riebler, S. H. Sørbye, D. P. Simpson, and H. Rue. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25(4):1145–1165, 2016.

B. Rosner, R. J. Glynn, and M. T. Lee. Wilcoxon signed rank test for paired comparisons of clustered data. *Biometrics*, 62(1):185–192, 2006. ISSN 0006-341X.

N. S. Rost, A. Bottle, J. Lee, M. Randall, S. Middleton, L. Shaw, V. Thijs, G. J. E. Rinkel, and T. M. Hemmen. Stroke severity is a crucial predictor of outcome: An international prospective validation study. *Journal of the American Heart Association*, 5(1):e002433, 2016. doi: 10.1161/JAHA.115.002433. URL `https://www.ahajournals.org/doi/abs/10.1161/JAHA.115.002433`.

M. Roy-O'Reilly and L. D. McCullough. Age and sex are critical factors in ischemic stroke pathology. *Endocrinology*, 159(8):3120–3131, 2018. doi: https://doi.org/10.1210/en.2018-00465.

H. Rue and L. Held. *Gaussian Markov random fields : theory and applications*, volume 104 of *Monographs on statistics and applied probability*. Chapman Hall/CRC, Boca Raton, Fla, 2005. ISBN 1584884320.

H. Rue, S. Martino, and N. 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.

H. Rue, A. Riebler, S. H. Sørbye, J. B. Illian, D. P. Simpson, and F. K. Lindgren. Bayesian computing with INLA: A review. *Annual review of statistics and its application*, 4(1):395–421, 2017. ISSN 2326-8298.

M. M. Rymer. Hemorrhagic stroke: intracerebral hemorrhage. *Missouri Medicine*, 108(1):50–54, 2011.

S. J. Sarker, P. U. Heuschmann, I. Burger, C. D. A. Wolfe, A. G. Rudd, N. C. Smeeton, and A. M. Toschke. Predictors of survival after haemorrhagic stroke in a multi-ethnic population: the south London stroke register (SLSR). 79(3):260–265, 2008. ISSN 0022-3050.

D. P. Simpson, H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye. Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1):1–28, 2017.

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

S. H. Sørbye and H. Rue. Penalised complexity priors for stationary autoregressive processes. *Journal of time series analysis*, 38(6):923–935, 2017. ISSN 0143-9782.

The Organization for Economic Cooperation and Development (OECD). Mortality following ischaemic stroke. URL `https://www.oecd-ilibrary.org/sites/a489af86-en/index.html?itemId=/content/component/a489af86-en`.

J. Van Niekerk, E. Krainski, D. Rustand, and H. Rue. A new avenue for Bayesian inference with INLA. *Computational statistics data analysis*, 181:107692, 2023. ISSN 0167-9473.