Spatio-temporal areal data modelling: COVID-19 applications and outlier detection for big data

Robin Muegge

School of Mathematics and Statistics

University of Glasgow

A thesis submitted for the degree of

Doctor of Philosophy

November 2024

Declaration

I, Robin Muegge, declare that this thesis titled, "Spatio-temporal areal data modelling: COVID-19 applications and outlier detection for big data" and the work presented in it are my own. I confirm that where I have consulted the published work of others, this is always clearly attributed.

The work presented in Chapter 3 has been published in the Spatial and Spatio-temporal Epidemiology journal with the title "National lockdowns in England: The same restrictions for all, but do the impacts on COVID-19 mortality risks vary geographically?" (Muegge et al., 2023), and is jointly authored by Dr Nema Dean, Dr Eilidh Jack, and Prof Duncan Lee. I delivered a poster presentation on this work at the Royal Statistical Society International conference in Aberdeen and a topic contributed talk in a remote session at the GEOMED conference in Irvine (CA), USA in 2022.

The work presented in Chapter 4 has been published in the Journal of the Royal Statistical Society - Series A with the title "COVID-19 vaccine fatigue in Scotland: How do the trends in attrition rates for the second and third doses differ by age, sex, and council area?" (Muegge et al., 2024), and is jointly authored by Dr Eilidh Jack, Dr Nema Dean, and Prof Duncan Lee. I delivered a topic contributed talk on this work at the Royal Statistical Society International conference in Harrogate in 2023.

I delivered a topic contributed talk on the work presented in Chapter 5 at the GEOMED conference in Hasselt, Belgium in 2024.

Everything we see hides another thing, we always want to see what is hidden by what we see.

— René Magritte

Abstract

The COVID-19 pandemic has been the greatest challenge to global public health in the 21st century. The novel virus demanded scientific progress in several fields of research, from medical innovations to the development of political strategies that aimed to contain the spreading of the virus and protect the most vulnerable, often assisted by statistical analyses. The work presented in this thesis is a timely analysis of important public health aspects of COVID-19 in the UK, detecting overall trends and patterns in mortality risk after the three national lockdowns in England and identifying differences in COVID-19 vaccine attrition rates for the second and third doses by age group, sex, and council area in Scotland. The presented statistical analyses fit spatio-temporal areal data using generalised linear mixed effects models in a Bayesian hierarchical framework, where the correlated spatial random effects are assigned prior distributions from the class of conditional autoregressive (CAR) models. These models typically induce spatial smoothness in the inferred disease risk or prevalence surface, where strength in the estimation is borrowed from neighbouring observations, according to some neighbourhood structure. The spatial smoothness assumption is often accredited to Waldo R. Tobler, who said, "Everything is related to everything else, but near things are more related than distant things". However, the presented COVID-19 analyses suggest that the spatial smoothness assumption might not always hold for all areas. Hence, this thesis proposes a novel relative density-based outlier score (RDOS) for identifying potential singleton spatial outliers that violate the spatial smoothness assumption and a novel modified spatial smoothing model to remove the potential outliers' impact on the estimated disease prevalence surface. The following summarises the key findings from this

thesis. The study on the impact of national lockdowns on COVID-19 mortality risk in England shows that the risks increased drastically before the implementation of lockdowns 1 and 3 and decreased to pre-lockdown levels after ten and six weeks, respectively. Further, the study identifies areas with a higher peak risk during these lockdowns, detecting an urban/rural divide for lockdown 1 and an association between higher risk and the early spreading of the Alpha variant during lockdown 3. The study on COVID-19 vaccine attrition rates in Scotland identifies a strong association between age and attrition rates, where the odds in favour of attrition decrease smoothly with increasing age. The odds in favour of attrition tend to be overall higher for males than females and higher in the second transition (from doses 2 to 3) than the first (from doses 1 to 2). Lastly, a simulation study shows that the novel singleton spatial outlier detection method for areal data produces much better detection results than the commonly used local Moran's I statistic. Similarly, the modified smoothing model is shown to produce overall better prevalence estimates than a conventional smoothing model when the number of outliers is large or at least some outliers have a large magnitude, even when the identified outlier sets are sub-optimal. The proposed methods are combined in a two-stage modelling approach and applied in a motivating study on asthma prevalence at the lower super output area (LSOA) level in England, where potential singleton spatial outliers are identified, and the estimated risk surface obtained from the modified smoothing model is compared to that of a conventional smoothing model. The comparison shows that the prevalence estimates of the identified outliers and their neighbouring inliers differ noticeably between the two models, highlighting the importance of considering such potential singleton spatial outliers in the analysis of areal unit data.

Acknowledgements

Firstly, I would like to thank my PhD supervisors, Prof Duncan Lee, Dr Nema Dean, and Dr Eilidh Jack. You have been incredibly generous in your support and guidance, pushing me beyond what I thought I was capable of and making my time as a PhD student a remarkable experience. Thank you for your patience, invaluable insight, and support of my academic and non-academic goals!

I would like to thank my examiners, Prof Christel Faes and Dr Jennifer Gaskell for their dedication and insightful discussion of my PhD thesis.

I completed my PhD under the Maclaurin scholarship in a part-time lecturing position. I thank the University of Glasgow for trusting in my abilities and for the financial support throughout my PhD.

I would not have dared attempt a PhD without the encouragement I received during my Master's programme from the excellent group of lecturers at Portland State University. Especially, I would like to thank Assoc Prof Daniel Taylor Rodríguez, Prof Jong Sung Kim, Prof Robert Fountain, and Teaching Asst Prof Nadeeshani Jayasena for their mentorship, enthusiasm, and inspiration in a shaping time of my life.

In the fourth year of my PhD programme, I completed a research placement at the University of Wollongong in Australia. I thank Assoc Prof Andrew Zammit Mangion for inviting me to spend ten weeks as a visiting research student, his kindness and generosity in sharing his statistical knowledge and experience, and his continued communication upon my return to Glasgow. I will be forever grateful for the experiences I've had on this research placement and associated travel.

I would also like to thank all the other researchers who shared their

knowledge and experience with me during my time in Australia and New Zealand. I especially want to thank Assoc Prof Susanna Cramb from the Queensland University of Technology, Dr Anton Westveld from the Australian National University, Prof James Brown from the University of Technology Sydney, Dr Tingjin Chu from the University of Melbourne, and Prof Daniel Exeter from the University of Auckland for inviting me to visit their institutions and hold seminar talks.

I received funding for my research placement in Wollongong from several sources. I want to thank the following funding bodies: G-Research for awarding me a PhD student grant, the University of Glasgow's Graduate School for awarding me a Mobility Scholarship, the London Mathematical Society for awarding me an Early Career Research Travel Grant, and the School of Mathematics and Statistics for providing additional funding.

I had the opportunity to present my PhD research at several conferences. I thank the University of Glasgow for funding my attendance at the RSS conferences in Aberdeen and Harrogate and my remote attendance at the GEOMED conference in Irvine (CA), USA. I would also like to thank the GEOMED organising committee for awarding me a travel grant to attend the conference in Hasselt.

I am grateful to have been part of a supportive PhD cohort with whom I've made many lasting memories. I especially thank Stephen Villejo and Daniela Cuba for being my friends throughout the PhD.

Lastly, I would like to thank my family and those closest to me: my parents, Heike and Michael, my brother Lukas, and my partner Helen. Michael, imagining your joy in any accomplishment I achieve will remain a great motivation throughout my life. Heike, your inexhaustible love and support carries for two. Lukas, you have been a great friend and inspiration since day one. Helen, thank you for all your fun, patience, tolerance, fun again, and commitment - I am incredibly lucky to share this life with you!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

According to the UK's National Health Service (NHS, 2024), the key functions of *public health* are health protection (e.g., infectious disease control and action for clean air), improving people's health (e.g., reducing health inequalities, for example by helping people quit smoking), and ensuring that the health services are effective, efficient, and equally accessible. The importance of public health is reflected in the 17 goals of the United Nations 2030 Agenda for Sustainable Development (United Nations - Department of Economic and Social Affairs, 2015), which was set out in 2015 and includes the goals of promoting well-being for all at all ages (Goal 3) and reducing inequality within and among countries (Goal 10). Many of the specified targets are phrased in terms of population statistics that are to be reached by 2030. Statistical analyses can be used to check the progress in reaching these specific targets and, more generally, to evaluate the key functionalities of public health and identify possibilities for improvement.

Historically, some of the most substantial public health improvements in the UK were achieved during the Victorian period from 1837 to 1901 (for a brief overview, see the timeline provided by the National Portrait Gallery, 2024). One of the earliest public health milestones during that period was the First Public Health Act of 1848, which initiated local boards of health to (among other things) appoint medical officers and build sewers. The Victorian period featured some of the earliest statistical analyses of public health data. The most renowned example is from 1854, when John

Snow identified a hotspot of cholera cases in London and discovered that a water pump was a likely source of the spread of cholera (e.g., see Rothman, 2012). The geographical analysis that led to the discovery is considered the earliest application in *epidemiology*, which can be defined as "the study of the determinants of the incidence and prevalence of disease" (Alderson, 1983). While the umbrella of epidemiological research is wide, some of its other key aims include estimating the number of individuals with a disease (possibly for different groups of people) and identifying patterns and trends in how these numbers change over space and time.

Florence Nightingale, one of the main pioneers of public health during the Victorian period and a member of the *sanitary movement* (e.g., see Susser and Stein, 2009), produced a key statistical analysis of temporal trends in public health data, which she utilised to call for political action to initiate further sanitary improvements. Nightingale analysed mortality data from military field hospitals during the Crimean War between 1854 and 1856, where she was given the role of Superintendent of Female Nursing (Small, 2017). It was William Farr, the Superintendent of the Statistical Department of the Registrar-General's Office, who mentored Nightingale and brought to her attention the association between the sanitary improvements (relating to the cleanliness of the crowded barracks, e.g., by enabling ventilation and building a system to flush water through the sewers that ran under the building) carried out in Spring 1955 and the reduction in mortalities that followed. Figure 1.1 shows Nightingale's 'coxcomb' or 'rose diagram', which visualises the before mentioned reduction in preventable deaths from disease. Nightingale's and Farr's successful collaboration largely contributed to the Second Public Health Act of 1875 (Small, 2017), whose main contributions included the clearing of slums and the refitting of sanitisation of existing dwellings, which ultimately led to a considerable increase in the population's average life expectancy and substantially improved public health, overall.

Modern spatio-temporal areal data models enable complex statistical analyses of spatially aggregated public health data of possibly large scope, and Bayesian disease mapping, in particular, has become an important subject in spatio-temporal epidemiological research with substantial methodological and computational advances over the last 40 years (e.g., see MacNab, 2022). Disease mapping studies aim, among

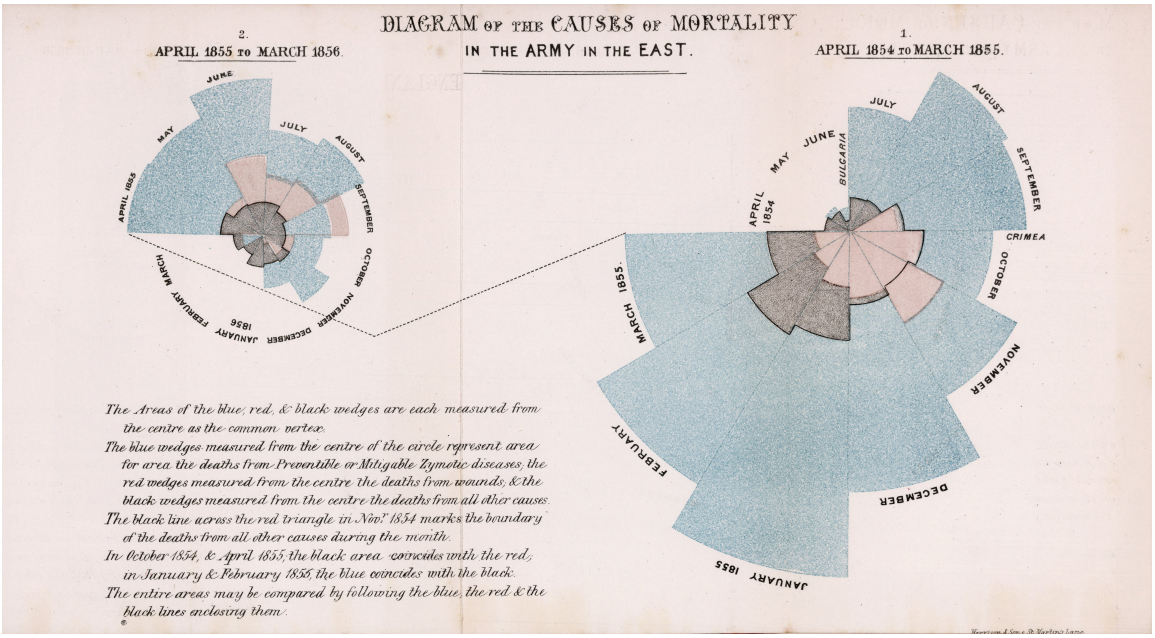


Figure 1.1: Diagram of the Causes of Mortality in the Army in the East by Florence Nightingale - https://www.davidrumsey.com/luna/servlet/s/h6xid2, Public Domain, https://commons.wikimedia.org/w/index.php?curid=1474443.

other things, to identify overall spatial patterns and temporal trends in disease risk, hotspots or clusters of high risk, factors associated with risk, and health inequalities. Many Bayesian disease mapping models belong to the class of conditional autoregressive (CAR) models, which tend to produce spatially smooth estimates of disease risk (or prevalence) and allow the identification of overall spatial patterns in the inferred risk (or prevalence) surface. Besag et al. (1991) proposed the earliest models from this class: the intrinsic conditional autoregressive (ICAR) model and convolution model (more commonly referred to as the BYM model, which is an acronym of the authors' names). The authors proposed the ICAR and BYM models specifically for modelling non-infectious diseases, and their work was motivated by cancer studies. Generally, much of the earlier disease mapping work was motivated by creating cancer risk maps (e.g., see MacNab, 2022). In contrast, in recent years, conditional autoregressive models have become highly relevant in analysing infectious diseases (e.g., see Tessema et al., 2023).

The work that led to the studies presented in this thesis commenced in October 2020, shortly after the first wave of the COVID-19 pandemic, which became the second leading cause of death worldwide in 2021 (World Health Organization, 2024). At the peak of the COVID-19 pandemic in the UK, the critical care facilities were saturated and could not admit any more patients (Venkatesan, 2024). Hence, the

pandemic challenged all key functions of public health: the spread of disease was no longer under control, critical medical operations had to be delayed or cancelled, and health services were not accessible to all. Given the pandemic's eminence in public health, the work presented in Chapters 3 and 4 provides a topical analysis of some of the most crucial public health aspects at the time of writing and is outlined below. Chapter 2 precedes these COVID-19 studies, reviewing relevant literature and methodology used in this thesis's three main chapters.

The work presented in Chapter 3 titled, "National lockdowns in England: The same restrictions for all, but do the impacts on COVID-19 mortality risks vary geographically?" presents a geographical analysis (at the local authority district level) of COVID-19 mortality risk after the three national lockdowns in England. The study's main aim is to investigate the spatio-temporal trends in COVID-19 mortality risk to identify geographical differences in the impact of lockdown, answering four main questions: 1. How long after the implementation of lockdown did mortality risks reduce at a national level, and did this vary by lockdown?; 2. How did the temporal trends in mortality risks differ by region in England?; 3. Which local authorities were exposed to the highest average risks in the weeks after lockdown?; and 4. Which local authorities shared similar temporal trends in mortality risks? The study duration is from 1st February 2020 to 14th May 2021, when no effective medical treatment for COVID-19 was available at a large scale. During that time, governments had to resort to *non-pharmaceutical* interventions, "measures to reduce transmission that did not depend on drugs, vaccines or other specific medical countermeasures" (Department of Health & Social Care, 2023), to slow down the spreading of the virus. These measures included social distancing, wearing face coverings, isolating infected individuals, and travel restrictions such as national lockdowns. The presented study appears to have been the first comprehensive investigation of spatio-temporal trends in COVID-19 mortality risks following the implementation of national lockdowns in England. Furthermore, in the review of studies on non-pharmaceutical interventions in the UK by Duval et al. (2024), the work presented in Chapter 3 was listed as one of only three ecological studies on measures to reduce the number of contacts evaluated for COVID-19 mortality (e.g., see the online resource

https://research.ukhsa.gov.uk/evidence-gap-map-npi-and-covid-19).

On 2nd December 2020, the first doses of COVID-19 vaccines were administered in the UK. By the summer of 2021, the COVID-19 vaccines had become available at a large scale and on 18th July 2021, everyone aged 18 years and older became eligible to receive a first dose of the vaccine (NHS, 2021). Following this progression of the pandemic, Chapter 4 presents another topical analysis of public health efforts. The study is titled, "COVID-19 vaccine fatigue in Scotland: How do the trends in attrition rates for the second and third doses differ by age, sex, and council area?" and provides an epidemiological analysis of the vaccination uptake by sex and age group across Scotland's council areas. The study's aim is to answer three main questions: 1. Are there any trends in attrition rates by age group, and does this vary by sex?; 2. Are there any spatial patterns in attrition rates by council area, and does this vary by sex?; and 3. How do these trends and patterns compare across the two transitions (from doses 1 to 2 and 2 to 3)? The presented study appears to have been the first to examine simultaneously the proportions of people who received the first dose of the vaccine but not a second dose and those who received the first and second doses but not the third (also known as *booster*). It proposes the term *vaccine fatigue* to refer to the delay in acceptance or refusal of subsequent doses to distinguish the findings from those of the more commonly studied *vaccine hesitancy*, which refers to a "delay in acceptance or refusal of vaccination despite availability of vaccination services" (MacDonald, 2015), i.e., the delay or refusal of any vaccine at all.

Public health data are often spatially autocorrelated, following Tobler's *First Law of Geography*, that "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). If an exploratory analysis supports the spatial smoothness assumption, one can fit a spatial smoothing model (e.g., a CAR model) to the data to account for the spatial autocorrelation. However, even when the data appear autocorrelated overall, the spatial smoothness assumption might not hold for a few areas deviating from the typical local behaviour. As a result, the unusual health outcomes in these areas are smoothed over, and they are likely to impact the inference of the otherwise smooth prevalence surface. In Chapter 4, the COVID-19 vaccine attrition rate for males in the council area Argyle and Bute is unusually

high in the transition from doses 1 to 2, and it appears that the spatial smoothness assumption does not hold for this observation. The discovery of the unusual attrition rate motivated the investigation of tools that can help identify such individual unusual areas in larger public health data (e.g., with thousands of areal units) and how one might account for such areas in a modified spatial smoothing model. These are the key motivating questions for the work presented in Chapter 5 titled, "When Tobler's First Law of Geography doesn't hold: Identifying spatially outlying observations to remove their impact on estimated disease prevalence surfaces".

The thesis concludes with Chapter 6, with a summary of the key findings of the studies presented in Chapters 3 to 5, and closes with a discussion of recognised limitations and proposes ideas for future work.

Chapter 2

Methodology and literature review

This chapter provides a review of the statistical methods and literature that will be applied and referred to throughout this thesis; further details are provided in the later chapters. The work presented in this thesis is motivated by applications in epidemiology, and Section 2.1 introduces the relevant disease mapping notation and methodology. The models presented in this thesis belong to the greater family of generalised linear models, which are reviewed in Section 2.2. The statistical inference is carried out in a Bayesian framework, and the key concepts of Bayesian statistics are reviewed in Section 2.3. The models are chosen based on the measures of model complexity and fit presented in Section 2.4. Types of spatio-temporal data and tools for analysing these data are presented in Section 2.5. Selected spatial and spatio-temporal models that can be applied to areal data are reviewed in Sections 2.6 and 2.7, respectively. Lastly, Section 2.8 reviews selected methods in cluster analysis.

2.1 Disease mapping

Public health analysts and epidemiologists seek to identify possible disease causes and assess the relationship between a disease and other factors that might impact its *incidence* (the number of new cases within a certain period) or *prevalence* (the proportion of people in a population having a disease), e.g., see Noordzij et al. (2010). Following the definition by Lawson et al. (1999), *disease mapping* refers to producing accurate maps of disease incidence to infer the true underlying disease distribution.

Some alternative names refer to the same subject, such as spatial epidemiology, environmental epidemiology, and small-area health studies (Lawson, 2018). Statistical methods are commonly applied to estimate the disease burden, investigate the spatial variation in disease risk and identify patterns in the inferred risk surface. This section reviews some general notation and methodology, but the existing literature on disease mapping is extensive; for a more thorough introduction, see, for example, the books by Waller and Carlin (2010) or Lawson (2018), or the recent review paper by MacNab (2022).

Generally, disease data are observed and aggregated for confidentiality and accuracy. For high-prevalence diseases, areas can be small enough to represent near-homogeneous communities, while for rare diseases, larger areas might be needed to capture the disease burden accurately. For example, https://statistics.gov.scot/atlas/resource?uri=http://statistics.gov.scot/id/statistical-geography/S92000003 shows some of the commonly used geographies in Scotland, such as Data Zones (6,976 areas), Intermediate Zones (1,279 areas), Council Areas (32 areas), and Health Boards (14 areas). Public health data are typically provided for such existing geographic boundaries, and the geographical scale might limit the availability of additional covariates or the conclusions that can be drawn from the statistical analysis. For small-area health studies, the relationship between disease prevalence and possible risk factors can be analysed but is subject to the availability of small-scale exposure data. Furthermore, as for any aggregated data, there is a possibility of *ecological bias* due to the within-area variability in exposures and confounders (Wakefield, 2008). Ecological bias is an even greater concern for large-scale analyses, as the within-area variability is anticipated to be greater for larger areas and could result in unreliable estimates of the relationships between disease prevalence and other factors.

Introducing some general concepts commonly applied in disease mapping, consider a public health analysis over a geographical region partitioned into K areas labelled $(1, \ldots, K)$ for which disease incidence counts $(Y_1, \ldots, Y_K)$ are observed. Analysing the raw counts does not allow for a fair comparison of the disease burden in the areas, as they do not consider the underlying population sizes or demographics (e.g.,

a larger number of disease counts might be expected in a larger population; similarly, if an older age is associated with a higher disease incidence, larger disease rates should appear more unusual in areas with younger populations). Hence, a more sophisticated statistic, the *standardised morbidity ratio* (SMR), is commonly computed as a proxy measure of an area's relative risk that accounts for the area's population size and some available demographics. Note that the SMR also denotes the *standardised mortality ratio*, computed from death counts rather than disease incidence counts, with otherwise analogous notation.

The SMR in area j, for $j = 1, \ldots, K$, is computed as

$$\hat{\theta}_j = \frac{Y_j}{e_j}, \tag{2.1}$$

where Y_j and e_j denote the area's observed and expected disease counts, respectively. For example, an SMR of $\hat{\theta}_j = 1.2$ implies that the observed count Y_j is 20% greater than the expected count e_j. The expected counts are computed via indirect standardisation. Generally, for d strata (e.g., demographic groups), the expected count in area j is computed as

$$e_j = \sum_{i=1}^{d} N_{ij} \times r_i, \tag{2.2}$$

where N_{ij} is the population size of stratum i in area j and r_i is the stratum-specific disease rate from a reference population. For example, the stratum-specific rates could be obtained at a national level, and the expected counts might be computed for the set of K areas in the country. Note that the areas' stratum-specific population sizes and the stratum-specific disease rates from the reference population must be known to compute the expected counts. These restrictions imply that the expected counts are often computed using only a few commonly reported demographics, such as age, sex, or ethnicity. As seen in Equation (2.2), the expected count e_j is computed assuming that the stratum-specific disease rates in area j are the same as in the reference population.

The SMR in Equation (2.1) is subject to substantial random variability whenever the expected count e_j is small, which can be the case when the area's population size is

small or consists largely of demographic groups with rare disease rates in the reference population. Statistical models are fitted to reduce some of this random variability in the SMR values. The observed counts $Y_1, \ldots, Y_K$ are commonly modelled using the Poisson distribution, such that for $j = 1, \ldots, K$,

$$Y_j \sim \text{Poisson}(\theta_j e_j). \tag{2.3}$$

Here, the parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$ denotes the true *relative risks* and is the target for inference. The interpretation of the *latent* (hidden or unobservable) relative risk θ_j is similar to that of the SMR, only that the relative risk measures how much the latent mean $E[Y_j]$ differs from the expected count e_j. For the assumed Poisson likelihood in Equation (2.3), the SMR's expectation is $E[\hat{\theta}_j] = E[Y_j/e_j] = \theta_j$; i.e., the observed SMR is an unbiased estimator of the true relative risk. It should be noted that the population sizes N_j in areas $j = 1, \ldots, K$ are finite and thus, one could argue that the binomial model would be more appropriate to fit the data, i.e., as $Y_j \sim \text{Binomial}(N_j, \alpha_j)$, where α_j denotes the latent disease incidence rate or prevalence in area j. However, for rare diseases, the observed counts tend to be much smaller than the population sizes, which legitimises using the Poisson distribution (e.g., see Waller and Carlin, 2010) for which the estimated relative risks have the intuitive interpretation provided above. When the binomial distribution is applied, the population demographics can still be considered when estimating the disease incidence rate or prevalence by fitting demographic-specific count data (e.g., see Chapter 4) or including the demography variables as covariates in the model.

The relative risk and disease prevalence are commonly modelled using a generalised linear model (see Section 2.2) in a Bayesian hierarchical framework (see Section 2.3). However, a generalised linear model might not fully explain the variability in the observed counts, which might be further explained by covariates that cannot be easily obtained and are hence not included in the model. A common assumption is that any covariate would be more similar for areas closer to each other, following Tobler's *First Law of Geography* (recall Chapter 1). If the spatial smoothness assumption holds for covariates that have not been included in the model but explain

some of the variability in the observed counts, the model can be improved by including spatially correlated random effects. Section 2.5 reviews methods for preliminary analysis of areal data, and Section 2.6 reviews a selection of spatial regression models for ecological analyses of areal data.

2.2 Generalised linear models

A generalised linear model (GLM) fits a linear regression to a variable's transformed mean. The method was originally proposed by Nelder and Wedderburn (1972), and the following review is based on the later work by McCullagh and Nelder (1989) and Wood (2017).

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_K)$ denote a vector of random variables that are assumed to have generated the observed data. The general structure of a GLM is

$$\begin{aligned} Y_j &\sim f(y_j|\mu_j), \\ g(\mu_j) &= \boldsymbol{x}_j\boldsymbol{\beta}, \end{aligned} \tag{2.4}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$ is a vector of the expected values of the random variables $\boldsymbol{Y}$ for the observed covariate values, i.e., $\mu_j = E[Y_j|\boldsymbol{x}_j = (x_{j1}, \ldots, x_{jp})]$, g is a smooth monotonic 'link function', $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is a vector of p unknown regression coefficients that are the target for inference, and $\boldsymbol{X}$ is a matrix of dimension $K \times p$ such that the j-th row contains the observed covariate values $\boldsymbol{x}_j$ corresponding to the observation Y_j. The variables Y_j are assumed to be conditionally independent (given the observed covariate values) and to follow some exponential family distribution $f(\cdot)$. That is, the probability density or probability mass function of the random variable Y_j can be written as

$$f_\theta(y_j) = \exp\{[y_j\theta - b(\theta)]/a(\phi) + c(y_j, \phi)\}, \tag{2.5}$$

where a, b and c are arbitrary functions, ϕ is a scale parameter, and θ is known as the canonical parameter of the distribution. The generalised linear model in Equation (2.4) takes on the form of a multiple linear regression model (e.g., see Mont-

gomery et al., 2012) when the variables Y_j are conditionally independent (given the observed covariate values) and identically Gaussian distributed, for $j = 1, \ldots, K$, and the link function $g(\cdot)$ is the identity function. In that case, $Y_j \sim \mathrm{N}(\mu_j = \boldsymbol{x}_j\boldsymbol{\beta}, \sigma^2)$ for $j = 1, \ldots, K$. Examples of distributions from the exponential family other than Gaussian include the Poisson, binomial, gamma, and beta distributions. Note that the generalised linear model in Equation (2.4) can easily be extended to a generalised linear mixed model by adding random effects. This thesis features generalised linear mixed models for the Poisson distribution in Chapter 3 and the binomial distribution in Chapters 4 and 5. The parameters in these models are estimated in a Bayesian framework, and the following section reviews the fundamental concepts of Bayesian statistics.

2.3 Bayesian statistics

Section 2.1 showed an example where the observed counts $(Y_1, \ldots, Y_K)$ are modelled using the Poisson distribution with means $(\theta_1 e_1, \ldots, \theta_K e_K)$. The model parameters are the relative risks $(\theta_1, \ldots, \theta_K)$, and the goal is to estimate these unknown parameter values. In frequentist statistics, these parameters are assumed to be fixed but unknown, and a common estimate is chosen to maximise the likelihood (hence called the *maximum likelihood estimator*, for short MLE). In Bayesian statistics, the parameters in the model are treated as random variables that are assigned prior distributions, which reflect what is known about the parameters before seeing the data. Considering the prior knowledge and assuming a data likelihood model allows deriving the posterior distribution of the parameters under consideration of the observed data. A brief introduction to these methods is presented below. For further information, see, for example, Gelman et al. (2013).

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ denote a parameter vector. Assume $\boldsymbol{y} = (y_1, \ldots, y_K)$ are K observations with probability density or probability mass function $f(\boldsymbol{y}|\boldsymbol{\theta})$. Assigning a prior density $f(\boldsymbol{\theta})$ to the parameter vector allows computing the posterior density of $\boldsymbol{\theta}$ given the observed data $\boldsymbol{y}$, which can be derived using Bayes' Theorem (Bayes,

1763) as

$$f(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\boldsymbol{y})}, \tag{2.6}$$

where $f(\boldsymbol{y}|\boldsymbol{\theta})$ is the joint data likelihood and $f(\boldsymbol{y})$ is a normalising constant. The joint data likelihood is interpreted as a function of the parameter $\boldsymbol{\theta}$ for fixed observations $\boldsymbol{y}$ and is commonly denoted using the alternative notation $f(\boldsymbol{y}|\boldsymbol{\theta}) = L(\boldsymbol{\theta}|\boldsymbol{y})$. The normalising constant $f(\boldsymbol{y})$ is free of $\boldsymbol{\theta}$; hence, it is not needed for inference on the parameter vector, and so the unnormalised posterior density of $\boldsymbol{\theta}$ given the data $\boldsymbol{y}$ can instead be written proportional to the product of the joint data likelihood and the prior density of the parameter vector, as

$$f(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}).$$

The posterior density of $\boldsymbol{\theta}$ (up to proportionality) can be used to identify ranges of values that are most likely and lead to a probabilistic interpretation. For example, when the range of values (a, b) contains $c\%$ of the posterior density of a parameter θ for some $0 \leq c \leq 100$, then there is a $c\%$ chance that the true parameter value is contained in the range (a, b). Posterior intervals can be computed either directly from the derived posterior distribution (when available in closed form), as credible intervals from simulated random draws from the posterior distribution (see Section 2.3.2), or estimated using approximate methods (see Section 2.3.3). Similarly, point estimates such as the posterior mean, median, or mode can be used for interpretation and communication.

2.3.1 Prior distributions

The prior distribution assigned to each model parameter should reflect the researcher's knowledge before seeing the data. When no prior information is available, this is commonly reflected in a non-informative or weakly informative prior distribution. For example, if the parameter θ denotes a probability, the parameter can be assigned a non-informative uniform prior distribution $\theta \sim \text{Unif}(0, 1)$. Here, the prior density

$f(\theta) = 1$ is the same for all $0 < \theta < 1$, which means that all values are equally likely before seeing the data (hence the term *non-informative*). When θ is a real-valued parameter, one can assign a weakly informative prior distribution such as a Gaussian prior, e.g., $\theta \sim \mathrm{N}(\mu = 0, \sigma^2 = 10,000)$. Here, the values closer to the mean $\mu = 0$ have a slightly higher density than the values further away from the mean; however, the density is fairly flat due to the distribution's large variance (hence the term *weakly informative*). These example prior densities are *proper*, meaning they integrate to one. Sometimes, one might choose an *improper* prior distribution, for example, when the prior density is proportional to a constant over an infinite range of values. It should be noted that improper priors can lead to a proper posterior density, meaning they can be valid for inference.

A commonly used class of prior distributions are *conjugate* priors. A prior distribution is conjugate for a data likelihood if the prior and posterior densities are from the same distribution. A *natural conjugate* prior has the same functional form as the data likelihood. Many commonly used distributions are members of the exponential family of distributions (i.e., their probability mass or density function can be expressed in the form of Equation 2.5), and for all members of the exponential family such a natural prior distribution exists (Gelman et al., 2013). For example, the beta prior is conjugate for the binomial distribution, the gamma prior is conjugate for the Poisson distribution, and the inverse-gamma prior is conjugate for the Gaussian distribution with known mean and unknown variance.

The previous examples treated the parameters as random variables, and the prior distributions of these parameters were parameter-free. For example, the weakly-informative Gaussian distribution had a mean of $\mu = 0$ and a variance of $\sigma^2 = 10,000$. Instead of fixing these parameter values, one can assign a prior distribution to the parameters of the prior distribution. The prior distribution's parameters are called hyperparameters, and the resulting model is referred to as a *hierarchical* model. More information on these types of models can be found in Gelman et al. (2013), and Banerjee et al. (2014) review these models specifically for spatial data. In this thesis, hierarchical models are implemented for spatial and spatio-temporal data, and some selected spatial and spatio-temporal models are reviewed in Sections 2.6 and

2.7, respectively. In Chapters 4 and 5, penalised complexity priors are used and this class of prior distributions is reviewed next.

2.3.1.1 Penalised complexity priors

Simpson et al. (2017) propose a framework for building priors for a large number of Bayesian models and call these priors *penalised complexity* (PC) priors. The authors introduce these priors for additive models in which the model components have flexibility parameters that control the assumed structure of the latent effects. In addition to the general method presented in this section, the authors propose a new parameterisation of the classical BYM model (see Section 2.6.1) for disease mapping, which Riebler et al. (2016) call the BYM2 model. The BYM2 model is reviewed in Section 2.6.4 and applied in Chapters 4 and 5, where additional information is provided relating to the presented studies.

Generally, PC priors penalise departures from simplified base models and are defined using probability statements about the prior distribution's flexibility parameter. The base model is the simplest version of a flexible density with flexibility parameter ζ. For example, consider an independent random effects vector $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)$ with a Gaussian joint prior distribution; specifically, $\boldsymbol{\phi}|\zeta \sim \mathrm{N}(\boldsymbol{\mu} = \mathbf{0}, \zeta\boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix of dimension $K \times K$. Here, the flexibility parameter ζ is the variance parameter of the Gaussian distribution, and the base model is obtained for $\zeta = 0$, which reflects the prior belief that the random effects are all equal to zero.

The construction of PC priors is based on four principles. Firstly, a PC prior penalises deviations from the base model, so the simpler model is preferred unless the observed data suggest otherwise (referred to as *Occam's razor*). Secondly, the PC prior measures the model complexity using the Kullback-Leibler divergence (KLD, Kullback and Leibler, 1951), which is computed as

$$\mathrm{KLD}(f(\boldsymbol{\phi}|\zeta)||f(\boldsymbol{\phi}|\zeta = 0)) = \int f(\boldsymbol{\phi}|\zeta) \log\left(\frac{f(\boldsymbol{\phi}|\zeta)}{f(\boldsymbol{\phi}|\zeta = 0)}\right) d\boldsymbol{\phi}. \tag{2.7}$$

The KLD in Equation (2.7) is the expectation of the log difference between the flexible density with parameter ζ and the base model with $\zeta = 0$. The larger the KLD,

the more information is lost when replacing the flexible model with the base model. The KLD is asymmetric, and Simpson et al. (2017) suggest using the unidirectional distance measure of complexity $d(\zeta) = \sqrt{2\text{KLD}(f(\boldsymbol{\phi}|\zeta)||f(\boldsymbol{\phi}|\zeta = 0))}$ of the flexible model with parameter ζ compared to the base-model with $\zeta = 0$. Thirdly, the deviation from the simpler model is penalised using a constant decay rate (also referred to as *constant-rate penalisation*). Specifically, for some constant decay rate $0 < r < 1$, the prior density $f_d(\cdot)$ applied to the distance scale satisfies the memoryless property

$$\frac{f_d(d+\delta)}{f_d(d)} = r^\delta, \text{ for } \delta \geq 0. \tag{2.8}$$

Thus, the decay rate does not depend on the value of d; i.e., increasing the distance d by an extra δ results in the same decay r^δ, regardless of the value d. Moreover, Equation (2.8) implies that the prior mode is at $d = 0$ (the base-model), as $f_d(\delta)/f_d(0) = r^\delta < 1$, which implies that $f_d(\delta) < f_d(0)$ for any $\delta > 0$. Simpson et al. (2017) assign an exponential prior density to the distance scale, which satisfies the memoryless property; specifically, $f_d(d) = \lambda \exp(-\lambda d)$ so that Equation (2.8) results in

$$\frac{f_d(d+\delta)}{f_d(d)} = \frac{\lambda \exp(-\lambda(d+\delta))}{\lambda \exp(-\lambda d)} = \exp(-\lambda\delta),$$

so that $r = \exp(-\lambda)$ for $\lambda > 0$. Simpson et al. (2017) state that the prior density on the original space is

$$f(\zeta) = \lambda \exp(-\lambda d(\zeta)) \left| \frac{\partial}{\partial \zeta} d(\zeta) \right|,$$

which follows from the change of variables formula when $d(\zeta)$ is a monotonous function of $\zeta > 0$ and d^{-1} exists such that $\zeta = d^{-1}(d(\zeta))$ and d^{-1} has a continuous derivative on the range of values that $d(\zeta)$ can take on (Casella and Berger, 2002). The fourth and final principle is that the user should have an idea of a sensible size for the parameter of interest. The authors propose selecting λ by controlling the prior

mass in the tails via the probability statement

$$P(Q(\zeta) > U) = \alpha,$$

where $Q(\zeta)$ is an interpretable transformation of the flexibility parameter, U is a user-defined upper bound for a tail event and α is the probability of $Q(\zeta)$ exceeding the threshold U. Thereby, the probability statement determines how informative the PC prior is.

Further information on the PC prior can be found in Simpson et al. (2017), and Chapters 4 and 5 present the PC priors specific to the BYM2 models (see Section 2.6.4) implemented in these studies.

2.3.2 Inference using Markov chain Monte Carlo (MCMC) simulations

For simple examples, one can derive the posterior density from the data likelihood and the prior density, as seen in Equation (2.6) to draw inferences on the parameter of interest. However, in practice, deriving the posterior density is often intractable or even impossible. As a resolution, one can use Markov chain Monte Carlo (MCMC) methods to draw simulated values from the posterior density, subject to the algorithm's convergence. The first MCMC methods were developed in the early 1950s but only became the predominant method for Bayesian inference in the 1980s-1990s, when computing power increased drastically, and many new methods were proposed. For a history of MCMC methods, see, for example, Robert and Casella (2011). The review of selected MCMC methods presented here is based on the work of Gelman et al. (2013).

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ denote the parameter vector of interest, possibly divided into d subvectors such that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d)$, where $\boldsymbol{\theta}_i$ denotes the i-th subvector, for $i = 1, \ldots, d$. The goal is to sample from the joint posterior density $f(\boldsymbol{\theta}|\boldsymbol{y})$, where $\boldsymbol{y} = (y_1, \ldots, y_K)$ denotes the observed data. The MCMC algorithm is initialised with some starting values $\boldsymbol{\theta}^0$, which can be specified manually or drawn from the prior distribution. The values are updated iteratively, and in iteration t (for $t =$

$1, 2, \dots$), the parameter values $\boldsymbol{\theta}^t$ are simulated under consideration of the values from the previous iteration $(t-1)$ but are conditionally independent of the other previous values in the chain $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \dots, \boldsymbol{\theta}^{(t-2)}$; note that the partial dependence on the latest value and conditional independence from the earlier values in the chain is known as the *Markov property* (hence the name *Markov chain*). The chain consists of simulated draws from the parameter space, and the distribution of the sampled values is expected to become closer to the true but unknown posterior density $f(\boldsymbol{\theta}|\boldsymbol{y})$; i.e., the posterior density is the stationary distribution of the Markov process.

The starting values $\boldsymbol{\theta}^0$ could be far from the high-density region of the true posterior density. As a result, it could take a considerable number of iterations before the chain approaches its stationary distribution. Hence, one should use a *burn-in* (also known as *warm-up*) period from which the simulated values are discarded as they are suspected not to reflect the true posterior density. Furthermore, since the new parameter values are generated conditionally on the parameter values from the previous draw, the simulated parameter values in the chain will not be independent. Hence, one can *thin* the chain by considering only every k-th sampled parameter vector for some $k \in \mathbb{N}_{>0}$, which reduces the dependence between the simulated values in the chain. The resulting sample of parameter values can then be used for inference on the parameter of interest, e.g., to compute point estimates such as the posterior mean or median or probabilistic properties such as credible intervals or exceedance probabilities.

For some parameters, it is possible to derive the full conditional distribution (given the other parameter values) in closed form. In that case, one can use the Gibbs sampler reviewed in Section 2.3.2.1 to sample the parameter values directly from their full conditional posterior densities. When the full conditional distribution cannot be derived in closed form but up to proportionality, one can use the Metropolis-Hastings algorithm reviewed in Section 2.3.2.2 to sample from the posterior densities. For either method, one should carry out convergence checks on the simulated values to rule out any obvious patterns that might indicate the Markov process has not converged, implying that the simulated draws do not accurately represent the target posterior densities. Section 2.3.2.3 reviews the convergence checks considered in this

thesis.

2.3.2.1 Gibbs sampler

Geman and Geman (1984) propose the Gibbs sampler, which performs alternating conditional sampling when the full conditional distributions can be derived in closed form. The algorithm starts with initial parameter values $\boldsymbol{\theta}^0$. In the t-th iteration, the parameter subvectors are successively sampled from their full conditional distributions, given all the other subvectors, in the following d steps:

1. draw $\boldsymbol{\theta}_1^t$ from $f(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2^{(t-1)}, \boldsymbol{\theta}_3^{(t-1)}, \ldots, \boldsymbol{\theta}_d^{(t-1)}, \boldsymbol{y})$,
2. draw $\boldsymbol{\theta}_2^t$ from $f(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1^t, \boldsymbol{\theta}_3^{(t-1)}, \ldots, \boldsymbol{\theta}_d^{(t-1)}, \boldsymbol{y})$,

   ...

d. draw $\boldsymbol{\theta}_d^t$ from $f(\boldsymbol{\theta}_d|\boldsymbol{\theta}_1^t, \boldsymbol{\theta}_2^t, \ldots, \boldsymbol{\theta}_{d-1}^t, \boldsymbol{y})$.

The above steps are completed for a specified number of iterations, and the simulated parameter values are used for inference. Convergence checks (see Section 2.3.2.3) should be carried out on the sampled parameter values to check that enough iterations have passed for the chain to have moved from the (possibly bad) initial values to exploring the parameter space.

2.3.2.2 Metropolis-Hastings algorithm

Hastings (1970) proposes the Metropolis-Hastings algorithm, which forms a generalisation of the method proposed by Metropolis et al. (1953). The algorithm allows simulating from a parameter's posterior density when its full conditional distribution cannot be derived in closed form. For the Metropolis-Hastings algorithm, a jumping distribution J_t is used to sample proposal values for the parameter vector. The algorithm again starts with initial parameter values $\boldsymbol{\theta}^0$. In iteration t, the following steps are completed for each of the d subvectors, here presented for the i-th subvector $\boldsymbol{\theta}_i$:

(a) Generate a proposal parameter subvector $\boldsymbol{\theta}_i^*$ from the jumping distribution $J_t\left(\boldsymbol{\theta}_i^* \mid \boldsymbol{\theta}_i^{(t-1)}\right)$, which considers the previous parameter subvector values $\boldsymbol{\theta}_i^{(t-1)}$.

(b) Compute the ratio

$$r = \frac{f(\boldsymbol{\theta}_i^*|\boldsymbol{y}) \Big/ J_t\left(\boldsymbol{\theta}_i^* \mid \boldsymbol{\theta}_i^{(t-1)}\right)}{f\left(\boldsymbol{\theta}_i^{(t-1)} \mid \boldsymbol{y}\right) \Big/ J_t\left(\boldsymbol{\theta}_i^{(t-1)} \mid \boldsymbol{\theta}_i^*\right)}. \tag{2.9}$$

(c) Set

$$\boldsymbol{\theta}_i^t = \begin{cases} \boldsymbol{\theta}_i^*, \text{ with probability } \min(r, 1), \\ \boldsymbol{\theta}_i^{(t-1)}, \text{ otherwise.} \end{cases}$$

Here, the target distribution $f(\boldsymbol{\theta}_i|\boldsymbol{y})$ must only be known up to proportionality, as any normalising constants cancel out since the target distribution is contained in both the numerator and denominator of the ratio r in Equation (2.9). When the jumping distribution J_t is symmetric so that $J_t\left(\boldsymbol{\theta}_i^* \mid \boldsymbol{\theta}_i^{(t-1)}\right) = J_t\left(\boldsymbol{\theta}_i^{(t-1)} \mid \boldsymbol{\theta}_i^*\right)$, the Metropolis-Hastings algorithm simplifies to the Metropolis algorithm, where the ratio in Equation (2.9) simplifies to $r = f(\boldsymbol{\theta}_i^*|\boldsymbol{y}) \Big/ f\left(\boldsymbol{\theta}_i^{(t-1)} \mid \boldsymbol{y}\right)$.

Depending on the specification of the jumping distribution, the obtained proposals could tend to be too far away from the previous values (when the jumps are too big, meaning that many of them could be rejected) or too close to the previous values (when the jumps are too small, meaning that the parameter space is not explored efficiently). In either of these two cases, the obtained simulated values might not accurately represent the parameter space. Hence, the jumping distribution can be tuned to produce jumps that effectively explore the parameter space. The study presented in Chapter 3 uses the `R` package `CARBayesST` (Lee et al., 2018), where the jumping distributions are automatically tuned to have acceptance rates between 40-50% for scalar parameter updates and between 20-40% for vector parameters.

2.3.2.3 Convergence checks

When using MCMC methods to infer properties of the parameter's posterior density from the sampled values, the quality of the inference will depend on how well the simulated values represent the parameter space. Thus, the inference relies on the

MCMC algorithm converging to the stationary distribution (the parameter's posterior density, which is the target for inference) before saving the simulated draws for inference. Hence, one should check that there are no obvious indications that the algorithm did not converge or does not explore the parameter space effectively.

One way to check possible convergence violations is by analysing the trace plots, i.e., plotting the sampled parameter values in the order of their position in the Markov chain. The trace plots should show random patterns across the range of parameter values and no clear trend. For example, the plot might show a clear trend for an early proportion of the iterations (e.g., increasing before changing into a random pattern over a consistent range of values), which suggests that the burn-in period was too short and that at least some samples were saved before the Markov chain converged to its stationary distribution.

Geweke (1992) proposes a convergence diagnostic that can be used (additionally to the visual checks) to assess possible violations of convergence. For $i = 1, \ldots, p$, let $\theta_i^{(t)}$ denote the simulated value of the i-th parameter in iteration t of the MCMC algorithm. Let s denote the total number of simulated values (post-burn-in and after thinning) for each parameter. Geweke's convergence diagnostic analyses the first and last part of the saved simulations from the Markov chain of each parameter in $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ and tests if the means of the posterior densities that these samples estimate can be assumed equal. The key idea is that if both samples were drawn from the stationary distribution (i.e., if the chain converged before saving any simulated draws), the means of the posterior densities from which these two samples are drawn should be equal. Geweke's convergence diagnostic computes the sample means from the two sections at the start and end of the chain and divides this difference by its estimated standard error. Hence, under the null hypothesis that the chain converged before saving any simulated values, Geweke's diagnostic asymptotically follows a standard normal distribution.

Let s_A denote the number of simulations considered from the start of the saved simulated values, and let s_B denote the number of simulations considered from the end of that chain. For the i-th parameter θ_i, let $\boldsymbol{\theta}_i^A = (\theta_i^{(1)}, \ldots, \theta_i^{(s_A)})$ denote the first s_A simulated values, and let $s^* = s - s_B + 1$ so that $\boldsymbol{\theta}_i^B = (\theta_i^{(s^*)}, \ldots, \theta_i^{(s)})$ denotes the

last s_B simulated values. Geweke's convergence diagnostic for the simulated values of the parameter θ_i is computed as

$$C_i = \frac{\bar{\theta}_i^A - \bar{\theta}_i^B}{\text{e.s.e.}\left(\bar{\theta}_i^A - \bar{\theta}_i^B\right)}, \tag{2.10}$$

where $\bar{\theta}_i^A = \frac{1}{s_A}\sum_{t=1}^{s_A}\theta_i^{(t)}$ and $\bar{\theta}_i^B = \frac{1}{s_B}\sum_{t=s^*}^{s}\theta_i^{(t)}$ denote the sample means and the denominator is the estimated standard error (e.s.e.) of their difference. Under the null hypothesis that the samples $\boldsymbol{\theta}_i^A$ and $\boldsymbol{\theta}_i^B$ are drawn from the stationary distribution, Geweke's diagnostic C_i is asymptotically normal distributed. Geweke (1992) uses consistent spectral density estimates for $\bar{\theta}_i^A$ and $\bar{\theta}_i^B$ to estimate the standard error. The study presented in Chapter 3 uses the `R` package `CARBayesST` (Lee et al., 2018) for model fitting, where the Geweke diagnostic is computed using the `geweke.diag()` function from the `coda` package (Plummer et al., 2005). Here, the spectral densities at zero are estimated by fitting autoregressive models (e.g., see Takalo et al., 2005) to the two sequences $\boldsymbol{\theta}_i^A$ and $\boldsymbol{\theta}_i^B$, and the code can be viewed in the package's publicly available GitHub repository at https://github.com/cran/coda.

Under the null hypothesis that the MCMC chain has converged, Geweke's diagnostic from Equation (2.10) asymptotically follows a standard normal distribution. Hence, for a significance level of 0.05, one should reject the null hypothesis if Geweke's convergence diagnostic takes on a value outside the 95% highest density interval of the standard normal distributions, i.e., if the statistic takes on any value below -1.96 or above 1.96. As a rule of thumb, a Geweke diagnostic between -2 and 2 does not indicate violations of convergence. It should be noted that Geweke's diagnostic can only be used to test for violations of the convergence (the null hypothesis is that the chain has converged), but the test cannot be used to conclude that the chain has converged. Hence, one should also assess convergence by looking at the trace plot of the saved simulated draws from the chain. Additional checks can be considered and for further discussion, see, for example, Section 11.4 of Gelman et al. (2013).

2.3.3 Inference using approximate methods

Markov chain Monte Carlo (MCMC) methods (see Section 2.3.2) allow accurate estimation of the posterior densities if the Markov chains have run sufficiently long so that the simulated draws are from the stationary distributions of interest. However, especially for more complex models and increasingly large datasets, the MCMC methods can be slow or even intractable. As a resolution, one can use approximate methods for Bayesian inference. A popular method with convenient implementation is provided by the `R-INLA` package, which is named after its initial foundational method proposed by Rue et al. (2009), using integrated nested Laplace approximation (INLA) for latent Gaussian models (i.e., models that assign Gaussian priors to the unobservable components).

The `R-INLA` package was developed with a focus on models that contain random effect vectors modelled by a joint prior distribution that has the Markov property: a random effect's full conditional prior density only depends directly on its neighbouring random effects; e.g., a temporal random effect's full conditional prior density might only depend directly on the effect from the preceding time point and a spatial random effect's full conditional prior density might only depend directly on the spatially neighbouring effects, according to some neighbourhood structure. A probability measure whose conditional distributions define such a neighbourhood structure is referred to as a *Markov random field* (Cressie, 1993); when the probability measure is Gaussian, it is referred to as a *Gaussian Markov random field* (GMRF, Rue and Held, 2005). Spatial and spatio-temporal models for areal data commonly contain random effects that are modelled using conditional autoregressive (CAR) priors (e.g., see the models presented in Sections 2.6 and 2.7), and these CAR models are conditional specifications of GMRFs. Hence, the `R-INLA` package can be used for inference in disease mapping applications that use CAR models, and the package is used for inference in the studies presented in Chapters 4 and 5.

It should be noted that, despite the package's name, the default method for inference in the `R-INLA` package is no longer integrated nested Laplace approximation. The original method was replaced by the modern formulation proposed by van Niek-

erk et al. (2023), which uses a variational Bayes correction to improve the lower-level Gaussian approximation. The modified version became the new default in `R-INLA` on 22nd November, 2022 (as documented under NEWS in `R-INLA`). Since the work in Chapters 4 and 5 was carried out after that date, the results presented in these chapters were obtained using the modern formulation presented in Section 2.3.3.1.

2.3.3.1 Gaussian approximations with variational Bayes correction

This section presents an approximate method for Bayesian inference that van Niekerk et al. (2023) propose and is (at the time of writing) the default method in the `R-INLA` package.

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_K)$ denote the random variables that are assumed to have generated the observations $\boldsymbol{y} = (y_1, \ldots, y_K)$ with density function $\pi(y|\boldsymbol{\mathcal{X}}, \boldsymbol{\theta})$, where $\boldsymbol{\mathcal{X}}$ denotes a latent field and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ a vector of hyperparameters. The linear predictors $\boldsymbol{\eta} = g(E[\boldsymbol{Y}])$ are fitted as a latent Gaussian model

$$\boldsymbol{\eta} = \beta_0 \mathbf{1} + \boldsymbol{X}\boldsymbol{\beta} + \sum_{l=1}^{L} f^l(\boldsymbol{u}_l),$$

where β_0 denotes an intercept term, $\mathbf{1}$ a vector of ones of length K, $\boldsymbol{X}$ and $\boldsymbol{U}$ denote covariate matrices where the j-th rows correspond to the j-th observation y_j, and $\{f^l : l = 1, \ldots, L\}$ are unknown functions of the covariate matrix $\boldsymbol{U}$. The target of inference is the latent field $\boldsymbol{\mathcal{X}} = (\beta_0, \boldsymbol{\beta}, \boldsymbol{f})$, which forms a vector of length m. The prior distribution of the latent field is $\boldsymbol{\mathcal{X}}|\boldsymbol{\theta} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{Q}_\pi^{-1}(\boldsymbol{\theta}))$, where $\boldsymbol{Q}_\pi^{-1}$ denotes a covariance matrix. Note that the hyperparameter vector can be written as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_1$ denotes the parameters of the likelihood and $\boldsymbol{\theta}_2$ the hyperparameters of the latent field. The following method approximates the marginal posteriors $\pi(\theta_j|\boldsymbol{y})$ and $\pi(\mathcal{X}_j|\boldsymbol{y})$.

One can express the generalised linear model as $\boldsymbol{\eta} = \boldsymbol{A}\boldsymbol{\mathcal{X}}$, where $\boldsymbol{A}$ denotes a sparse design matrix so that the joint posterior density of the latent field and hyperparameters can be written as

$$\pi(\boldsymbol{\mathcal{X}}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\mathcal{X}}|\boldsymbol{\theta}) \prod_{j=1}^{K} \pi(y_j|(\boldsymbol{A}\boldsymbol{\mathcal{X}})_j, \boldsymbol{\theta}).$$

The joint posterior density of the hyperparameter values is estimated using the Laplace approximation

$$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) = \left.\frac{\pi(\boldsymbol{\mathcal{X}}, \boldsymbol{\theta}|\boldsymbol{y})}{\pi_G(\boldsymbol{\mathcal{X}}|\boldsymbol{\theta}, \boldsymbol{y})}\right|_{\boldsymbol{\mathcal{X}}=\boldsymbol{\mu}(\boldsymbol{\theta})},$$

where $\pi_G(\boldsymbol{\mathcal{X}}|\boldsymbol{\theta}, \boldsymbol{y})$ denotes a Gaussian approximation of the full conditional posterior density $\pi(\boldsymbol{\mathcal{X}}|\boldsymbol{\theta}, \boldsymbol{y})$, and $\boldsymbol{\mu}(\boldsymbol{\theta})$ denotes its mode. The joint Gaussian approximation at the mode is obtained from a second-order expansion (see van Niekerk et al., 2023 for details). The approximation is of the form $\boldsymbol{\mathcal{X}}|\boldsymbol{\theta}, \boldsymbol{y} \sim \mathrm{N}\big(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{Q}_{\mathcal{X}}^{-1}(\boldsymbol{\theta})\big)$, where $\boldsymbol{Q}_{\mathcal{X}}^{-1}(\boldsymbol{\theta}) = \boldsymbol{Q}(\boldsymbol{\theta}) + \boldsymbol{A}^T\boldsymbol{D}\boldsymbol{A}$ with $\boldsymbol{D}$ denoting a diagonal matrix that depends on $\boldsymbol{\theta}$. The marginal full posterior distributions can be estimated from the joint Gaussian approximation as

$$\mathcal{X}_j|\boldsymbol{\theta}, \boldsymbol{y} \sim \mathrm{N}\Big((\boldsymbol{\mu}(\boldsymbol{\theta}))_j, \big(\boldsymbol{Q}_{\mathcal{X}}^{-1}(\boldsymbol{\theta})\big)_{jj}\Big).$$

Finally, the marginal posterior density of $\mathcal{X}_j$ is estimated by integrating out $\boldsymbol{\theta}$ using n integration points $\boldsymbol{\theta}^{(k)}$ with area weights δ_k, so that

$$\tilde{\pi}(\mathcal{X}_j|\boldsymbol{y}) \approx \sum_{k=1}^{n} \pi_G(\mathcal{X}_j|\boldsymbol{\theta}^{(k)}, \boldsymbol{y})\tilde{\pi}(\boldsymbol{\theta}^{(k)}|\boldsymbol{y})\delta_k.$$

Similarly, the marginal posterior density of θ_j can be estimated by integrating out $\boldsymbol{\theta}_{-j} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_p)$ from the estimated joint posterior density $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$. Note that the estimated marginal posterior densities of the linear predictors, $\tilde{\pi}(\eta_j|\boldsymbol{y})$ (for $j = 1, \ldots, K$), are calculated after the estimates $\tilde{\pi}(\mathcal{X}_j|\boldsymbol{y})$ and $\tilde{\pi}(\theta_j|\boldsymbol{y})$ are obtained. For details on how these are computed, see van Niekerk et al. (2023).

The posterior means of the linear predictors and the latent field can be improved using a variational Bayes correction, as proposed by van Niekerk and Rue (2024). The variational function is of the form

$$E_{q(\boldsymbol{\mathcal{X}}|\boldsymbol{y})}[-\log(l(\boldsymbol{\mathcal{X}}|\boldsymbol{y}))] + \mathrm{KLD}[q(\boldsymbol{\mathcal{X}}|\boldsymbol{y})||\pi(\boldsymbol{\mathcal{X}})],$$

where $q(\cdot)$ is a member of the variational class, $\pi(\cdot)$ is the prior density, $l(\cdot)$ is the

likelihood function, and KLD denotes the Kullback-Leibler divergence (e.g., see Equation 2.7). The approximation $\boldsymbol{\mathcal{X}}|\boldsymbol{\theta}, \boldsymbol{y} \sim \mathrm{N}\big(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{Q}_{\mathcal{X}}^{-1}(\boldsymbol{\theta})\big)$ uses the mean $\boldsymbol{\mu}(\boldsymbol{\theta})$, which is calculated in the Gaussian approximation at the mode of $\pi(\boldsymbol{\mathcal{X}}|\boldsymbol{y}, \boldsymbol{\theta})$. Using a variational Bayes correction, the improved mean takes the form

$$\boldsymbol{\mu}^*(\boldsymbol{\theta}) = \boldsymbol{\mu}(\boldsymbol{\theta}) + \boldsymbol{M}\boldsymbol{\lambda},$$

where $\boldsymbol{M}$ is a matrix that propagates the correction made to p nodes to the rest of the latent field (see van Niekerk and Rue, 2024 for details), and $\boldsymbol{\lambda}$ denotes the explicit corrections. The explicit corrections are solved for iteratively as

$$\underset{\lambda}{\operatorname{argmin}}\left(E_{\mathcal{X}|\boldsymbol{y},\boldsymbol{\theta}\sim\mathrm{N}\left(\boldsymbol{\mu}(\boldsymbol{\theta})+\boldsymbol{M}\boldsymbol{\lambda}, \boldsymbol{Q}_{\mathcal{X}}^{-1}(\boldsymbol{\theta})\right)}[-\log(l(\boldsymbol{\mathcal{X}}|\boldsymbol{y}))] + \frac{1}{2}(\boldsymbol{\mu}(\boldsymbol{\theta}) + \boldsymbol{M}\boldsymbol{\lambda})^T \boldsymbol{Q}(\boldsymbol{\theta})(\boldsymbol{\mu}(\boldsymbol{\theta}) + \boldsymbol{M}\boldsymbol{\lambda}) \right),$$

where $\log(l(\boldsymbol{\mathcal{X}}|\boldsymbol{y})) = \sum_{j=1}^{K} \log(\pi(y_j|(\boldsymbol{A}\boldsymbol{\mathcal{X}})_j))$, and the expected log-likelihood is approximated using a second-order Taylor series expansion around $\boldsymbol{\lambda} = \boldsymbol{0}$ (see van Niekerk et al., 2023 for details). Lastly, the marginal posterior densities of the linear predictors are approximated under consideration of the improved mean as

$$\begin{aligned}
\eta_j|\boldsymbol{\theta}, \boldsymbol{y} &\sim \mathrm{N}(\mu_j(\boldsymbol{\theta}), \sigma_j^2(\boldsymbol{\theta})), \\
\mu_j(\boldsymbol{\theta}) &= (\boldsymbol{A}\boldsymbol{\mu}^*(\boldsymbol{\theta}))_j, \\
\tilde{\pi}(\eta_j|\boldsymbol{y}) &\approx \sum_{k=1}^{n} \pi_G(\eta_j|\boldsymbol{\theta}^{(k)}, \boldsymbol{y})\tilde{\pi}(\boldsymbol{\theta}^{(k)}|\boldsymbol{y})\delta_k.
\end{aligned}$$

2.4 Model comparison

If models were compared by how well they fit the observed data, one could end up using an over-parameterised model that fits the data too closely, meaning that it is not a good model for unseen data or predictions of future observations. Instead, the model's predictive accuracy can be used to compare candidate models or decide which variables to include in a model. Information criteria are measures of predictive accuracy, and this thesis considers the deviance information criterion (DIC) presented in Section 2.4.2 and the Watanabe-Akaike information criterion (WAIC, also referred

to as *widely applicable information criterion*) presented in Section 2.4.3. First, the simpler Akaike information criterion (AIC) is introduced in Section 2.4.1. Note that for each information criterion, a smaller value indicates a better predictive accuracy. The review presented here is based on the work of Gelman et al. (2013).

In the following subsections, let $\boldsymbol{y} = (y_1, \ldots, y_K)$ denote a vector of K observations and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ a vector of p model parameters.

2.4.1 AIC

The Akaike information criterion (AIC, Akaike, 1974) is computed as

$$\text{AIC} = -2\log(p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_{\text{MLE}})) + 2p,$$

where $\log(p(\boldsymbol{y}|\boldsymbol{\theta}))$ is the log-likelihood (also referred to as *log predictive density*), $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ is the maximum likelihood estimate of $\boldsymbol{\theta}$, and p is the number of parameters estimated in the model. The smaller the AIC, the better the model's predictive accuracy. Adding the term $2p$ is a correction for how much the fitting of p parameters will increase predictive accuracy and is based on the asymptotic normal distribution. Hence, the AIC should only be used for linear models with flat (i.e., non-informative) priors, while more advanced information criteria are needed for hierarchical models or when informative priors are used.

2.4.2 DIC

The deviance information criterion (DIC, Spiegelhalter et al., 2002) is an alternative to the AIC that uses a data-based bias correction to penalise the number of parameters included in the model and replaces the MLE $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ from the AIC with the posterior mean $\hat{\boldsymbol{\theta}}_{\text{Bayes}} = E[\boldsymbol{\theta}|\boldsymbol{y}]$. The criterion is defined as

$$\text{DIC} = -2\log(p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_{\text{Bayes}})) + 2p_{\text{DIC}},$$

where the effective number of parameters (functioning as a bias correction) is computed from simulated values $\boldsymbol{\theta}^s$ for $s = 1, \ldots, S$ as

$$p_{\text{DIC}} = 2\left(\log(p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_{\text{Bayes}})) - \frac{1}{S}\sum_{s=1}^{S}\log(p(\boldsymbol{y}|\boldsymbol{\theta}^s))\right).$$

2.4.3 WAIC

The Watanabe-Akaike criterion (WAIC, Watanabe, 2010) is another information criterion that uses data-based bias correction. Gelman et al. (2013) define the criterion as

$$\text{WAIC} = -2\sum_{j=1}^{K}\log(p(y_j|\hat{\boldsymbol{\theta}})) + 2p_{\text{WAIC}}$$

and recommend letting $p_{\text{WAIC}} = \sum_{j=1}^{K}\text{Var}(\log(p(y_j|\boldsymbol{\theta})))$. In practice, the log pointwise predictive density is computed as $\sum_{j=1}^{K}\log(p(y_j|\hat{\boldsymbol{\theta}})) \approx \sum_{j=1}^{K}\log(\frac{1}{S}\sum_{s=1}^{S}p(y_j|\boldsymbol{\theta}^s))$ and the effective number of parameters is computed as

$$p_{\text{WAIC}} \approx \sum_{j=1}^{K}\text{Var}_{s=1}^{S}(\log(p(y_j|\boldsymbol{\theta}^s))),$$

where $\text{Var}_{s=1}^{S}(a(s)) = \frac{1}{S-1}\sum_{s=1}^{S}(a(s) - \bar{a})^2$, for $\bar{a} = \frac{1}{S}\sum_{s=1}^{S}a(s)$ and where $\boldsymbol{\theta}^s$ are simulated values. The provided definition of the WAIC criterion is used by the `R-INLA` package (see Section 2.3.3), which is used for inference in the studies in Chapters 4 and 5, where different models are compared using the WAIC.

2.5 Spatio-temporal data

Waldo R. Tobler's *First Law of Geography* states that "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970), where *near things* refers to spatial proximity. For example, environmental risk factors (e.g., air pollution) might change smoothly across the map, or lifestyle choices in communities living close by and interacting with each other might be more similar than in communities that live further apart. The same assumption can be applied to

the temporal domain. For example, disease risks might change gradually as the underlying population remains mostly the same in consecutive periods. Such similarities can be identified in observed data as spatial or temporal autocorrelation.

The correlation between two variables can be measured using Pearson's correlation coefficient (Pearson, 1896). For two observation vectors $\boldsymbol{x} = (x_1, \ldots, x_K)$ and $\boldsymbol{y} = (y_1, \ldots, y_K)$, the correlation coefficient is computed as

$$r = \frac{\sum_{j=1}^{K}(x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{j=1}^{K}(x_j - \bar{x})^2 \sum_{j=1}^{K}(y_j - \bar{y})^2}},$$

where $\bar{x} = \frac{1}{K}\sum_{j=1}^{K} x_j$ and $\bar{y} = \frac{1}{K}\sum_{j=1}^{K} y_j$ denote the sample means. Thus, the correlation coefficient is computed as the sample covariance between the two variables, divided by the product of their sample standard deviations. Sections 2.5.1 and 2.5.2 review methods to test for significant spatial and temporal autocorrelation, respectively. Rather than computing correlations between variables, these measures indicate how similar observations from the same variable are over space or time.

Note that different types of spatial data exist. Spatial statistics methodology commonly refers to geostatistical data, lattice data, or point patterns (e.g., see Cressie, 1993). This thesis focuses on areal data, where the geographical structure can be described using irregular lattices. Thus, the methods reviewed for investigating spatial autocorrelation in Section 2.5.1 and the selection of spatial and spatio-temporal models reviewed in Sections 2.6 and 2.7 apply to irregular lattice data and areal data, specifically.

2.5.1 Spatial autocorrelation

Section 2.5.1.1 reviews Moran's I statistic and a formal hypothesis test for overall spatial autocorrelation. Section 2.5.1.2 reviews local Moran's I statistic, which can be used for identifying differences in local spatial autocorrelation.

2.5.1.1 Moran's I

Consider spatial data with K areal units, for which the values $\boldsymbol{y} = (y_1, \ldots, y_K)$ are observed. Let $\boldsymbol{W}$ denote a neighbourhood matrix where $w_{ij} = 1$ if areas i and j

are neighbours and $w_{ij} = 0$, otherwise. For example, neighbourhoods can be defined using a border-sharing rule so that $w_{ij} = 1$ if areas i and j share a border, or using a k-nearest neighbour rule so that $w_{ij} = 1$ if area j is among the k nearest neighbours of area i, where proximity is measured geographically and commonly uses the Euclidean distance between the areas' centroids. By construction, $w_{jj} = 0$ for all j. The global Moran's I statistic (Moran, 1950) is computed as

$$I = \frac{K \sum_{i=1}^{K} \sum_{j=1}^{K} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{(\sum_{i=1}^{K} \sum_{j=1}^{K} w_{ij}) \sum_{j=1}^{K} (y_j - \bar{y})^2},$$

where $\bar{y} = \frac{1}{K} \sum_{j=1}^{K} y_j$ is the mean of the observed values. Moran's I statistic takes on values from -1 to 1, and positive values indicate positive spatial autocorrelation in $\boldsymbol{y}$ (which means that neighbouring areas tend to be more similar).

One can compute Moran's I statistic and run a permutation test to evaluate if the data are significantly positively spatially autocorrelation. Specifically, one permutes the original data $\boldsymbol{y}$ to obtain new data $\tilde{\boldsymbol{y}}$ and computes Moran's I statistic for the permuted observations. Computing Moran's I statistic for a large number of such permuted datasets, one can compute the proportion of these Moran's I statistics that are larger than the Moran's I statistic of the original data (hence, providing a probability of observing a stronger autocorrelation than in the original data, by chance). If the proportion is smaller than some significance level (e.g., $\alpha = 0.05$), then one can conclude that the original data $\boldsymbol{y}$ are significantly positively spatially autocorrelated. In that case, a spatial smoothing model might produce better estimates of the underlying risk surface than a non-spatial model that estimates the random effects independently.

Generally, one could include covariates in a generalised linear model (see Section 2.2) to measure the strength of the association between these explanatory variables and the response. When the covariates explain the spatial autocorrelation in the observed data, a non-spatial model that contains these explanatory variables might fit the data well. However, when the covariates cannot explain some of the spatial variations in the observed data, the standard errors of the coefficient estimates of ordinary least squares models will likely be too small (e.g., see Ripley, 1988 or

Cressie, 1993). In that case, a spatial model can account for the spatial correlation and increase the variance of the coefficient estimates. Section 2.6 reviews a selection of some commonly used models for areal data.

2.5.1.2 Local Moran's I

Local Moran's I statistic (Anselin, 1995) is computed for area j as

$$I_j = \frac{K(y_j - \bar{y})}{\sum_{i=1}^{K}(y_i - \bar{y})^2} \sum_{i=1}^{K} w_{ij}(y_i - \bar{y}), \tag{2.11}$$

so the global Moran's I statistic is proportional to the sum of the local Moran's I statistics and can be computed as $I = \frac{1}{\sum_{i=1}^{K}\sum_{j=1}^{K} w_{ij}} \sum_{j=1}^{K} I_j$. The local Moran's I statistic I_j takes on a negative value when $(y_j - \bar{y})$ and $\sum_{i=1}^{K} w_{ij}(y_i - \bar{y})$ have different signs; i.e. when y_j is either a high value surrounded by neighbouring values that tend to be low or a low value surrounded by neighbouring values that tend to be high, where high and low are in comparison to the overall average $\bar{y}$. Anselin (1995) claims that for the local Moran's I statistic, "a positive value indicates spatial clustering of similar values (either high or low)". It should be noted that not only spatial clusters of similar values can result in a positive local Moran's I value. For example, if an area's observation is much larger than the overall mean and the areas in its neighbourhood tend to be slightly larger than the overall mean, the local Moran's I value will be positive, despite the values' dissimilarity (only that they tend to be on the same side of the overall mean). Individual areas that deviate from the typical local behaviour are the focus of Chapter 5, where local Moran's I statistic is considered a competitor to a novel singleton spatial outlier detection method for identifying such unusual areas.

2.5.2 Temporal autocorrelation

Temporal autocorrelation measures how similar observations of a single variable are over time. The following notation is taken from Chatfield (2003).

For observations $\boldsymbol{y} = (y_1, \ldots, y_N)$ from N discrete time points, the temporal autocorrelation between observations from successive time points is computed as the

coefficient

$$r_1 = \frac{\sum_{t=1}^{N-1}(y_t - \bar{y})(y_{t+1} - \bar{y})}{\sum_{t=1}^{N}(y_t - \bar{y})^2}.$$

Instead of considering the correlation between observations at successive time points, one could consider the correlation observed at a greater lag. Generally, the temporal autocorrelation coefficient at lag l for some $l \in \mathbb{N}_{>0}$ is computed as

$$r_l = \frac{\sum_{t=1}^{N-l}(y_t - \bar{y})(y_{t+l} - \bar{y})}{\sum_{t=1}^{N}(y_t - \bar{y})^2}. \tag{2.12}$$

One can test for temporal autocorrelation by making distributional assumptions under the null hypothesis that the data are temporally independent. However, the data might only be autocorrelated at specific lags. Rather than performing separate hypothesis tests for temporal autocorrelation at different lags, one could consider a simultaneous hypothesis test of temporal autocorrelation over a range of lags from one up to m for some $m \in \mathbb{N}_{>0}$. One available method is the Ljung-Box test, which is presented next.

2.5.2.1 Ljung-Box test

The Ljung-Box test is a hypothesis test of temporal autocorrelation for lags l from one up to m for some $m \in \mathbb{N}_{>0}$. Ljung and Box (1978) proposed the method for identifying a lack-of-fit in time series models, where the test statistic is computed from the temporal autocorrelation coefficients of the residuals from the fitted model. If the time series model fits the data adequately (under the null-hypothesis), the residuals should appear as white noise, resulting in an asymptotic distribution of the test statistic. Instead of evaluating the fit of a time series model, the exploratory data analysis of Chapter 3 computes the Ljung-Box statistic from the temporal autocorrelation coefficients of the observed outcomes to check if the data are temporally autocorrelated to warrant using a temporal model. The following review is based on Harvey (1993), who describes applying the Ljung-Box test to the raw data.

Consider the sample autocorrelation coefficients r_l from Equation (2.12). When the observations are independent, the sample autocorrelation coefficients are asymp-

totically normal with mean zero and variance $1/N$. Thus, one can perform a hypothesis test where the null hypothesis is that the autocorrelation coefficients are independent (note that this assumption is stronger than a lack of temporal autocorrelation). Based on the asymptotic property under the null hypothesis, one can compute the Ljung-Box statistic

$$\tilde{Q}_m(r) = N(N+2)\sum_{l=1}^{m}\frac{r_l^2}{N-l},$$

which is a weighted sum of the temporal sample autocorrelation coefficients up to lag m. Under the null-hypothesis of temporal independence, $\tilde{Q}_m(r)$ approximately follows a Chi-square distribution with m degrees of freedom, χ^2_m. Thus, if the statistic $\tilde{Q}_m(r)$ exceeds the quantile of the distribution χ^2_m specified by the test's significance level, one can reject the hypothesis of temporal independence and conclude that there is significant evidence of temporal autocorrelation, which warrants using a temporal model.

2.6 Spatial models for areal data

This section presents a selection of spatial models that account for the spatial autocorrelation (see Section 2.5) in the observed data. Specifically, these models are designed for areal data and are commonly used in disease mapping applications (see Section 2.1). The presented models belong to the class of generalised linear mixed models (see Section 2.2) and are formulated in a Bayesian hierarchical framework (see Section 2.3). For a more extensive review of spatial models commonly used in disease mapping applications, see, for example, Lee (2011) or MacNab (2022).

Spatial models require the specification of the geographical structure to account for the data's spatial autocorrelation. For areal data, the spatial models typically use neighbourhood matrices to capture partial autocorrelation between areas defined as

neighbours. In the general case, the neighbourhood matrix $\boldsymbol{W}$ has cells

$$w_{ij} = \begin{cases} 1, & \text{if areas } i \text{ and } j \text{ are neighbours,} \\ 0, & \text{otherwise.} \end{cases} \tag{2.13}$$

Here, the neighbourhood matrix assigns equal weights to neighbouring areas. For example, one could define the matrix so that $w_{ij} = 1$ if areas i and j share a border, or if one area is among the k-nearest neighbours of the other, for some integer $k > 0$. Generally, rather than using a binary neighbourhood matrix, one could assign weights to indicate distances between areas or similarities in covariates observed in these areas. Limitations of the different measures are discussed, for example, by Gleditsch and Ward (2001). In this thesis, only binary neighbourhood matrices are applied to avoid making the models overly complex.

For the following models, assume data $\boldsymbol{Y} = (Y_1, \ldots, Y_K)$ are observed in areas $(1, \ldots, K)$, where each observation Y_j is assumed to have been generated by some data likelihood with mean μ_j. The mean is estimated using a generalised linear mixed model, which extends the generalised linear model from Equation (2.4) by including additive random effects. The following sections present each model's random effects specifications.

2.6.1 BYM and the intrinsic CAR model

Besag et al. (1991) propose a model that contains additive spatial effects. The model is commonly referred to as the BYM (Besag-York-Mollie) model and is of the form

$$g(\mu_j) = \beta_0 + \boldsymbol{x}_j\boldsymbol{\beta} + \phi_j + \psi_j, \text{ for } j = 1, \ldots, K,$$

where β_0 denotes the overall mean, $\boldsymbol{x}_j = (x_{j1}, \ldots, x_{jp})$ is a vector of covariates observed for area j with corresponding regression coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$, ϕ_j is a structured spatial random effect, and ψ_j is an unstructured spatial random effect for area j. The unstructured effects $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_K)$ are assumed to be independent and identically normal distributed; i.e. $\psi_j \sim N(0, \sigma^2)$ for all j, where σ^2 is a variance

parameter. Thus, their joint distribution is $\boldsymbol{\psi} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I})$, where $\boldsymbol{I}$ is an identity matrix of dimension $K \times K$.

The prior distribution of the structured spatial random effects $\boldsymbol{\phi}$ is constructed around a binary adjacency matrix $\boldsymbol{W}$ as defined in Equation (2.13). The structured random effects are assumed to be independent of the unstructured random effects. Their joint distribution is $\boldsymbol{\phi} \sim \mathrm{N}(\mathbf{0}, \tau^2 \boldsymbol{Q}(\boldsymbol{W})^-)$, where the precision matrix is defined as $\boldsymbol{Q}(\boldsymbol{W}) = \mathrm{diag}(\boldsymbol{W}\mathbf{1}) - \boldsymbol{W}$ with $\mathbf{1}$ denoting a vector of ones of length K and $\boldsymbol{Q}(\boldsymbol{W})^-$ denoting the generalised inverse (Penrose, 1955). The term $\mathrm{diag}(\boldsymbol{W}\mathbf{1})$ denotes a diagonal matrix where the j-th diagonal element is the number of neighbouring areas of area j. The full conditional prior density of ϕ_j can then be shown to take on the form

$$f(\phi_j | \boldsymbol{\phi}_{-j}, \tau^2) \propto \exp\left\{ -\frac{1}{2\tau^2} \sum_{i \sim j} (\phi_j - \phi_i)^2 \right\},$$

where $\boldsymbol{\phi_{-j}} = (\phi_1, \ldots, \phi_{j-1}, \phi_{j+1}, \ldots, \phi_K)$ is a vector that contains all spatial random effects except for ϕ_j, and $i \sim j$ indicates that areas i and j are neighbours (i.e., $i \sim j$ if and only if $w_{ij} = w_{ji} = 1$). The full conditional density can be rewritten as

$$\phi_j | \boldsymbol{\phi}_{-j}, \tau^2 \sim N\left(\frac{\sum_{i=1}^K w_{ij} \phi_i}{\sum_{i=1}^K w_{ij}}, \frac{\tau^2}{\sum_{i=1}^K w_{ij}} \right),$$

and this distribution is known as intrinsic conditional autoregressive (ICAR) prior distribution. The full conditional mean of the spatial random effect ϕ_j is the average of the random effects from the neighbouring areas of area j. All random effects share a common variance parameter τ^2, and the variance term in the full conditional density is smaller for areas with a larger number of neighbouring areas, reflecting the assumption of strong spatial autocorrelation in estimating the random effects. It should be noted that the distribution is only well defined if for each j, $w_{ij} = 1$ for at least one $i \neq j$; i.e., each area must have at least one neighbour. As a closing remark on the BYM model, note that only the sum of the random effects ϕ_j and ψ_j is identifiable. For further discussion on how this property might impact the statistical inference, see, for example, Eberly and Carlin (2000).

2.6.2 Leroux CAR model

Leroux et al. (2000) propose a generalisation of the ICAR prior distribution that allows for weaker spatial autocorrelation. The generalised linear mixed model takes the form

$$g(\mu_j) = \beta_0 + \boldsymbol{x}_j\boldsymbol{\beta} + \phi_j, \text{ for } j = 1, \ldots, K,$$
$$\boldsymbol{\phi} \sim N\big(\mathbf{0}, \tau^2\boldsymbol{Q}(\rho, \boldsymbol{W})^{-1}\big),$$

with precision matrix $\boldsymbol{Q}(\rho, \boldsymbol{W}) = \rho(\text{diag}(\boldsymbol{W}\mathbf{1}) - \boldsymbol{W}) + (1-\rho)\boldsymbol{I}$, where $\boldsymbol{I}$ denotes a $K \times K$ identity matrix and $\mathbf{1}$ is a vector of ones of length K. The full conditional distribution of the spatial random effect ϕ_j for area j, given the spatial random effects of all other areas, can then be expressed as

$$\phi_j|\boldsymbol{\phi}_{-j} \sim N\left(\frac{\rho\sum_{i=1}^{K} w_{ij}\phi_i}{\rho\sum_{i=1}^{K} w_{ij} + 1 - \rho}, \frac{\tau^2}{\rho\sum_{i=1}^{K} w_{ij} + 1 - \rho}\right),$$

where $\boldsymbol{\phi}_{-j} = (\phi_1, \ldots, \phi_{j-1}, \phi_{j+1}, \ldots, \phi_K)$ is the vector of spatial random effects of all areas except for area j, ρ controls the degree of spatial autocorrelation, and τ^2 is a variance parameter. Recall that in the ICAR prior (see Section 2.6.1), the full conditional mean of ϕ_j is the average of the random effects in the neighbouring areas of area j. The full conditional mean of ϕ_j under the Leroux CAR prior is a weighted average of the random effects in the neighbouring areas of area j, where the weight is determined by the correlation parameter ρ. For $\rho = 0$, the conditional mean is zero, and for $\rho = 1$, the model is identical to the ICAR prior distribution and implies strong spatial autocorrelation. For $0 < \rho < 1$, the full conditional variance term again becomes smaller as the number of neighbouring areas increases. The Leroux CAR prior distribution is well defined as long as each row $i = 1, \ldots, K$ in the neighbourhood matrix $\boldsymbol{W}$ has at least one cell $w_{ij} = 1$ (i.e., each area must have at least one neighbouring area).

2.6.3 Dean model

Dean et al. (2001) propose an alternative formulation that allows for weaker spatial

autocorrelation. Here, the generalised linear mixed model takes the form

$$g(\mu_j) = \beta_0 + \boldsymbol{x}_j\boldsymbol{\beta} + u_j, \text{ for } j = 1, \ldots, K,$$
$$\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{D}_u),$$

with covariance matrix $\boldsymbol{D}_u = \sigma_u^2(\lambda\boldsymbol{Q}^- + (1-\lambda)\boldsymbol{I}_u)$. The matrix $\boldsymbol{Q}$ is the same as for the ICAR model (see Section 2.6.1), $\boldsymbol{Q}^-$ again denotes its generalised inverse, and $\boldsymbol{I}_u$ denotes an identity matrix of size $K \times K$. The parameter λ determines how much of the overall variability in the random effects $\boldsymbol{u}$ is captured as a spatial or unstructured variation. For $\lambda = 1$, the variability in $\boldsymbol{u}$ is completely explained by the spatial structure and for $\lambda = 0$, the random effects $\boldsymbol{u}$ are independent and identically distributed. Hence, the model is similar to the Leroux CAR model (see Section 2.6.2) in allowing for weaker spatial autocorrelation. However, rather than formulating the precision matrix as a combination of a spatial and an unstructured precision matrix, the Dean model decomposes the covariance matrix as a combination of a spatial and an unstructured covariance matrix.

2.6.4 BYM2 model

Simpson et al. (2017) propose a modified version of the Dean model (see Section 2.6.3) where the spatial random effects are scaled. Specifically, Simpson et al. (2017) propose the model

$$g(\mu_j) = \beta_0 + \boldsymbol{x}_j\boldsymbol{\beta} + \frac{1}{\sqrt{\tau}}(\sqrt{1-\phi}v_j + \sqrt{\phi}u_j^*), \text{ for } j = 1, \ldots, K,$$

where $0 \leq \phi \leq 1$ is a mixing parameter, $\boldsymbol{u}^*$ are scaled spatial random effects, and $\boldsymbol{v}$ are unstructured random effects. Letting $\boldsymbol{b} = \frac{1}{\sqrt{\tau}}(\sqrt{1-\phi}\boldsymbol{v} + \sqrt{\phi}\boldsymbol{u}^*)$, the covariance matrix of the effects $\boldsymbol{b}$ is of the form $\frac{1}{\tau}((1-\phi)\boldsymbol{I} + \phi\boldsymbol{Q}_*^-)$, where $\boldsymbol{Q}_*$ denotes a scaled version of the precision matrix $\boldsymbol{Q}$ of the ICAR model (see Section 2.6.1) and $\boldsymbol{Q}_*^-$ denotes its generalised inverse.

Scaling the structured random effects allows for interpretability of the hyperprior distribution of the precision parameter τ and for transferability between applications

(Riebler et al., 2016) so that similar hyperpriors can be used for comparable applications with differing structure matrices. Following the methodology proposed by Sørbye and Rue (2014), the correlated random effects are standardised using the generalised variance $\sigma^2_{GV}(\boldsymbol{u})$, which is computed as the geometric mean of the marginal variances, i.e.

$$\sigma^2_{GV}(\boldsymbol{u}) = \exp\left(\frac{1}{K}\sum_{j=1}^{K}\log\left(\frac{1}{\tau_u}[\boldsymbol{Q}^-]_{jj}\right)\right),$$

where $\boldsymbol{Q}$ denotes the precision matrix. The precision matrix is scaled to $\boldsymbol{Q}^*$ such that $\sigma^2_{GV}(\boldsymbol{u}^*) = 1$ for $\tau_u = 1$.

2.7 Spatio-temporal models for areal data

Spatio-temporal models can be applied to capture spatial and temporal trends in data. They can be formulated in either a *separable* or *inseparable* form. If one part of the model accounts for spatial trends while another part accounts for temporal trends, then the model is called separable. The model is called inseparable if the spatial and temporal trends are captured simultaneously. Here, assume that the aim is to estimate a disease risk (or prevalence) θ_{jt} for areas $j = 1, ..., K$ in periods $t = 1, ..., N$, and this section reviews a selection of spatio-temporal models that can be used for inference.

2.7.1 Bernardinelli model

Bernardinelli et al. (1995) propose a model of the form

$$g(\theta_{jt}) = \beta_0 + \boldsymbol{x}_j\boldsymbol{\beta} + \phi_j + (\beta_1 + \delta_j)t,$$

where ϕ_j denotes a spatial random effect for area j, β_1 is a mean linear time trend over all areas, and δ_j denotes an interaction effect between time and area. The index t represents equally distant periods $t = 1, ..., N$ with the units depending on the context of the problem. If the time periods were not equally distant, then t could be replaced by t_i, where $i = 1, ..., N$ indicates the respective time period so that the

distances in time $\{t_i - t_{i-1}\}$ do not need to be the same for $i = 2, ..., N$. Note that to comply with identifiability, the effects must satisfy the restrictions $\sum_{j=1}^{K} \phi_j = 0$ and $\sum_{j=1}^{K} \delta_j = 0$.

Furthermore, it should be noted that the Bernardinelli model restricts both the main and area-specific time trends to be linear, an assumption that should be checked carefully via exploratory analysis of the data before applying this model.

2.7.2 Knorr-Held model

Knorr-Held (2000) extends the BYM model (see Section 2.6.1) by including two vectors of temporal effects $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, and a vector of spatio-temporal interaction effects $\boldsymbol{\delta}$. The model takes on the form

$$g(\theta_{jt}) = \beta_0 + \boldsymbol{x}_j\boldsymbol{\beta} + \phi_j + \psi_j + \alpha_t + \gamma_t + \delta_{jt},$$

where ϕ_j and ψ_j are structured and unstructured spatial random effects for area j, and α_t and γ_t are structured and unstructured temporal random effects for time period t, respectively. There are different options for the structure of the spatio-temporal random effects in $\boldsymbol{\delta}$ that will be discussed after the introduction of the random effect vectors $\boldsymbol{\phi}$, $\boldsymbol{\psi}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\gamma}$.

Each of the spatial or temporal random effect vectors follows a normal distribution with mean zero and precision matrix $\lambda\boldsymbol{S}$, where λ is an unknown scalar and $\boldsymbol{S}$ is a known structure matrix. That is, $\boldsymbol{\alpha} \sim \mathrm{N}\big(\mathbf{0}, (\lambda_\alpha \boldsymbol{S}_\alpha)^{-1}\big)$, $\boldsymbol{\gamma} \sim \mathrm{N}\big(\mathbf{0}, (\lambda_\gamma \boldsymbol{S}_\gamma)^{-1}\big)$, $\boldsymbol{\phi} \sim \mathrm{N}\big(\mathbf{0}, (\lambda_\phi \boldsymbol{S}_\phi)^{-1}\big)$, and $\boldsymbol{\psi} \sim \mathrm{N}\big(\mathbf{0}, (\lambda_\psi \boldsymbol{S}_\psi)^{-1}\big)$.

The prior distribution of the structured spatial effects $\boldsymbol{\phi}$ is an ICAR model (Section 2.6.1). That is, $\boldsymbol{\phi} \sim \mathrm{N}\big(\mathbf{0}, (\lambda_\phi \boldsymbol{S}_\phi)^{-1}\big)$, where $\boldsymbol{S}_\phi = \mathrm{diag}(\boldsymbol{W}\mathbf{1}) - \boldsymbol{W}$ and $\boldsymbol{W}$ denotes the neighbourhood matrix defined in Equation (2.13). The spatial random effects $\boldsymbol{\psi}$ are assumed independent and unstructured, so that $\boldsymbol{\psi} \sim \mathrm{N}\big(\mathbf{0}, (\lambda_\psi \boldsymbol{S}_\psi)^{-1}\big)$, where $\boldsymbol{S}_\psi = \boldsymbol{I}$ is an identity matrix of dimension $K \times K$.

The joint prior distribution of the structured temporal random effects $\boldsymbol{\alpha}$ is a

first-order random walk in time, such that

$$f(\boldsymbol{\alpha}|\lambda_\alpha) \propto \exp\left\{-\frac{\lambda_\alpha}{2}\sum_{t=2}^{N}(\alpha_t - \alpha_{t-1})^2\right\},$$

i.e. $\boldsymbol{\alpha} \sim \mathrm{N}\big(\mathbf{0}, (\lambda_\alpha \boldsymbol{S}_{\boldsymbol{\alpha}})^{-1}\big)$, with

$$\boldsymbol{S}_{\boldsymbol{\alpha}} = \begin{bmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{bmatrix}. \tag{2.14}$$

The full conditional distribution for α_t can be shown to be $\alpha_t|\boldsymbol{\alpha}_{-t}, \lambda_\alpha \sim \mathrm{N}(\mu_\alpha, \sigma_\alpha^2)$, where $\boldsymbol{\alpha}_{-t} = (\alpha_1, \ldots, \alpha_{t-1}, \alpha_{t+1}, \ldots, \alpha_N)$ is the vector of temporal effects for all time points except for t. The full conditional mean of each structured temporal random effect is determined by its neighbouring temporal effect(s), and the variance term of the full conditional distribution for those points that fall between two time periods is half of that of the random effects from the first ($t = 1$) and last ($t = N$) periods. The temporal random effects $\boldsymbol{\gamma}$ are assumed independent and unstructured, so that $\boldsymbol{\gamma} \sim \mathrm{N}\big(\mathbf{0}, (\lambda_\gamma \boldsymbol{S}_\gamma)^{-1}\big)$, where $\boldsymbol{S}_\gamma = \boldsymbol{I}$ is an identity matrix of dimension $N \times N$.

The structure matrix of the spatio-temporal random effect vector $\boldsymbol{\delta}$ is computed by taking the Kronecker product of the structure matrices from one of the spatial effects vectors and one of the temporal effects vectors. Generally, for two matrices $\boldsymbol{A}_{K\times L}$ and $\boldsymbol{B}_{M\times N}$, taking the Kronecker product (denoted $\otimes$) results in a $KM \times LN$ block matrix

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & \ldots & a_{1L}\boldsymbol{B} \\ \vdots & \vdots & \vdots \\ a_{K1}\boldsymbol{B} & \ldots & a_{KL}\boldsymbol{B} \end{bmatrix},$$

where a_{jt} denotes the entry in the j-th row and t-th column of matrix $\boldsymbol{A}$.

With two options for each spatial and temporal random effect vector (one structured and the other unstructured), there are four possible combinations to incorporate spatio-temporal effects in the model. Depending on the effects chosen, the spatio-temporal effects can be either *a priori* independent, have a structure in time but not in space, have a structure in space but not in time, or have a structure in both space and time that considers first and second order neighbours (i.e., spatial neighbours, temporal neighbours, spatial neighbours of temporal neighbours, or temporal neighbours of spatial neighbours). Any of the spatio-temporal random effects in a given period can be written as vectors $\boldsymbol{\delta_1}, \ldots, \boldsymbol{\delta_N}$, where $\boldsymbol{\delta_t} = (\delta_{1t}, \ldots, \delta_{Kt})$ is the vector of random effects for period t. These random effects have a normal joint prior distribution $\boldsymbol{\delta} \sim \mathrm{N}(\mathbf{0}, (\lambda_\delta \boldsymbol{S}_\delta)^{-1})$, where $\boldsymbol{S}_\delta = \boldsymbol{S}_{\text{time}} \otimes \boldsymbol{S}_{\text{space}}$ is of dimension $NK \times NK$. Here, $\boldsymbol{S}_{\text{time}}$ is either $\boldsymbol{S}_\alpha$ or $\boldsymbol{S}_\gamma$, and $\boldsymbol{S}_{\text{space}}$ is either $\boldsymbol{S}_\phi$ or $\boldsymbol{S}_\psi$. For example, spatio-temporal random effects that have a structure in both time and space can be obtained by letting $\boldsymbol{S}_\delta = \boldsymbol{S}_\alpha \otimes \boldsymbol{S}_\phi$.

As a limitation of the method, note that the Knorr-Held model is over-parameterised. For example, the structured and unstructured spatial random effects and the spatio-temporal random effects all account for the spatial trends in the data. Hence, it is unclear how much of the spatial trends should be picked up by each of these random effects. The same issue arises for the temporal trends in the data.

2.7.3 Ugarte model

Ugarte et al. (2012) provide a spatio-temporal model with CAR priors applied to the spatial, temporal, and spatio-temporal random effects. They model the risk (or prevalence) θ_{jt} as

$$g(\theta_{jt}) = \beta_0 + \boldsymbol{x}_j \boldsymbol{\beta} + \phi_j + \gamma_t + \delta_{jt},$$

where ϕ_j denotes the spatial random effect of area j, γ_t denotes a temporal random effect at time point t, and δ_{jt} denotes a spatio-temporal interaction effect for area j and time point t. The joint prior distributions for each of the vectors $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_N)$, and $\boldsymbol{\delta} = (\delta_{11}, \ldots, \delta_{K1}, \delta_{12}, \ldots, \delta_{K2}, \ldots, \delta_{1N}, \ldots, \delta_{KN})$ are then de-

fined as

$$\begin{aligned}
\boldsymbol{\phi} &\sim \mathrm{N}(\mathbf{0}, \sigma_s^2 \boldsymbol{D}_s); & \boldsymbol{D}_s &= (\lambda_s \boldsymbol{Q}_s + (1-\lambda_s)\boldsymbol{I}_s)^-, \\
\boldsymbol{\gamma} &\sim \mathrm{N}(\mathbf{0}, \sigma_t^2 \boldsymbol{D}_t); & \boldsymbol{D}_t &= \boldsymbol{Q}_t^-, \\
\boldsymbol{\delta} &\sim \mathrm{N}(\mathbf{0}, \sigma_{st}^2 \boldsymbol{D}_{st}); & \boldsymbol{D}_{st} &= \boldsymbol{Q}_t^- \otimes \boldsymbol{Q}_s^-,
\end{aligned}$$

where $\boldsymbol{Q}_s = \mathrm{diag}(\boldsymbol{W}\mathbf{1}) - \boldsymbol{W}$ with a neighbourhood matrix $\boldsymbol{W}$ (see Equation 2.13), $\boldsymbol{I}_s$ denotes an identity matrix of size $K \times K$, and $\boldsymbol{Q}_t = \mathrm{diag}(\boldsymbol{M}\mathbf{1}) - \boldsymbol{M}$, where $m_{ij} = 1$ if i and j are consecutive periods and 0, otherwise. Thus, $\boldsymbol{D}_s$ is the structure matrix of a Leroux CAR prior with λ_s as spatial autocorrelation parameter, and $\boldsymbol{D}_t$ is the structure matrix of an ICAR prior for temporal effects, which can be understood as a first-order random walk in time. That is, $\boldsymbol{D}_t$ is of the same form as $\boldsymbol{S}_\alpha$ in Equation (2.14). The structure matrix $\boldsymbol{D}_{st}$ for the spatio-temporal interaction effects is obtained by taking the Kronecker product between the structure matrices $\boldsymbol{Q}_t^-$ and $\boldsymbol{Q}_s^-$ of the temporal and spatial random effects.

Thus, the Ugarte model is similar to the Knorr-Held model in that it uses a spatio-temporal structure matrix that is created by taking the Kronecker product of the structure matrices of the spatial and temporal random effect vectors. A key difference between the models is that, unlike the Knorr-Held model, the Ugarte model does not contain any additional unstructured spatial or temporal random effects. However, the Ugarte model could be considered as over-parameterised as it is unclear to what extent the spatial random effects or the spatio-temporal random effects should account for the spatial trends in the data. The same holds for the temporal trends in the data. It should be noted that the Ugarte model contains only a single spatial random effects vector that allows for weaker autocorrelation (replacing the linear combination of a strongly positively spatially autocorrelated and independent random effects vector), while the single temporal random effects vector enforces strong temporal autocorrelation.

2.7.4 Rushworth model

Rushworth et al. (2014) propose a spatio-temporal model with non-separable spatio-

temporal random effects. Here, the risk (or prevalence) θ_{jt} is modelled as

$$g(\theta_{jt}) = \beta_0 + \boldsymbol{x}_j\boldsymbol{\beta} + \phi_{jt},$$

where ϕ_{jt} denotes a spatio-temporal random effect. The joint prior distribution of the spatio-temporal random effects vector $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N)$ is decomposed as

$$f(\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N) \sim f(\boldsymbol{\phi}_1) \prod_{t=2}^{N} f(\boldsymbol{\phi}_t | \boldsymbol{\phi}_{t-1}),$$

where $\boldsymbol{\phi}_t = (\phi_{1t}, \ldots, \phi_{Kt})$ denotes the vector of random effects for time period t.

The joint prior distribution of $f(\boldsymbol{\phi}_1)$ is a Leroux CAR model (see Section 2.6.2) that induces a spatial structure in the random effects from the first period. Thus, the full conditional distribution for the spatio-temporal random effect of area j at period 1 is given by

$$\phi_{j1} | \boldsymbol{\phi}_{-j1} \sim N\left(\frac{\rho \sum_{k=1}^{K} w_{jk}\phi_{k1}}{\rho \sum_{k=1}^{K} w_{jk} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{k=1}^{K} w_{jk} + 1 - \rho} \right),$$

where $\boldsymbol{W}$ is a neighbourhood matrix (see Equation 2.13) and ρ is a spatial correlation parameter.

The conditional prior distributions $f(\boldsymbol{\phi}_t | \boldsymbol{\phi}_{t-1})$, for $t = 2, \ldots, N$ capture the temporal autocorrelation in the random effects, and are given by

$$\boldsymbol{\phi}_t | \boldsymbol{\phi}_{t-1} \sim \mathrm{N}(\alpha\boldsymbol{\phi}_{t-1}, \tau^2 \boldsymbol{Q}(\rho, \boldsymbol{W})^{-1}), \text{ for } t = 2, \ldots, N,$$

where the precision matrix $\boldsymbol{Q}(\rho, \boldsymbol{W})$ is again that of the Leroux CAR prior (Section 2.6.2). The temporal autocorrelation is induced via the mean $\alpha\boldsymbol{\phi}_{t-1}$, where α is a temporal autocorrelation parameter; i.e. $\alpha = 0$ indicates temporal independence, while $\alpha = 1$ indicates strong temporal autocorrelation and turns this conditional part of the decomposition into a first-order random walk. The spatial autocorrelation is induced via the precision matrix determined by the neighbourhood structure captured in $\boldsymbol{W}$.

Note that this is a first order temporal autoregressive process (AR(1)). While

Rushworth et al. (2014) only provide this model, the implementation of the model in the function `ST.CARar()` in the `CARBayesST` package (Lee et al., 2018) also allows for a second order temporal autoregressive process (AR(2)). Chapter 3 shows an application that uses the AR(2) version of the model, and provides the additional methodology.

As a closing remark, the joint distribution of the random effects can be compared more easily to the Ugarte model (see Section 2.7.3) when expressed in the following form:

$$f(\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N) = f(\boldsymbol{\phi}_1) \prod_{t=2}^{N} f(\boldsymbol{\phi}_t | \boldsymbol{\phi}_{t-1})$$
$$\propto \exp\left\{ -\frac{1}{2} \left(\boldsymbol{\phi}_1^T \boldsymbol{Q}(\rho, \boldsymbol{W}) \boldsymbol{\phi}_1 + \sum_{t=2}^{N} (\boldsymbol{\phi}_t^T - \alpha \boldsymbol{\phi}_{t-1}^T) \boldsymbol{Q}(\rho, \boldsymbol{W}) (\boldsymbol{\phi}_t - \alpha \boldsymbol{\phi}_{t-1}) \right) / \tau^2 \right\}.$$

A key difference between the Rushworth model and the Ugarte model is how they account for the temporal structure: the Rushworth model uses an autoregressive prior of order one (AR(1)), and the Ugarte model uses a random walk prior of order one. Thus, the Rushworth model allows for weaker temporal autocorrelation, while the Ugarte model imposes strong temporal autocorrelation. Furthermore, the Rushworth model contains only one random effect per area and period. In contrast, the Ugarte model contains a spatial, temporal, and spatio-temporal random effect per area and period.

2.8 Cluster analysis

Clustering methods aim to split a collection of objects into subsets (i.e., groups or classes), which are referred to as "clusters". The objective is to group the objects so that objects in the same cluster have more similar attributes and objects that are not in the same cluster have more dissimilar attributes from one another. Thus, a cluster analysis requires some (dis-)similarity measure between the objects' attribute values. The standard clustering approaches can be divided into *hierarchical* and *non-hierarchical* clustering methods (e.g., see Giordani et al., 2020). Hierarchical methods produce a series of partitions that can be presented in a hierarchy tree where

clusters are merged or split as one moves along the tree's branches. Non-hierarchical methods, also known as *partitioning* methods, find the best partition (according to some loss function) for splitting the n objects into exactly k clusters. Thus, for non-hierarchical methods, one has to specify the number of clusters k before applying the method, and the clustering method will produce only one clustering with k clusters (as opposed to the hierarchy tree produced by the hierarchical methods). Selected non-hierarchical and hierarchical clustering methods are presented in Sections 2.8.1 and 2.8.2, respectively.

The work presented in this thesis applies the non-hierarchical clustering methods k-means and k-medoids, the latter of which is also known as *partitioning around medoids*. Section 2.8.1.1 reviews the k-means algorithm and Section 2.8.1.2 reviews the partitioning around medoids algorithm. Since one has to specify the number of clusters before applying these non-hierarchical clustering algorithms, one usually applies the methods for different numbers of clusters and decides which of the resulting clusterings appears to be the best partition of the set of objects. Some procedures for doing so are reviewed in Section 2.8.1.3. Section 2.8.1.4 reviews the Rand index, which can be used to measure the similarity between two clusterings. Lastly, Section 2.8.2 reviews single linkage agglomerative clustering, which is a hierarchical clustering method used in Chapter 3 to check for potential outliers before applying the k-means algorithm.

2.8.1 Non-hierarchical clustering (partitioning)

This section reviews selected non-hierarchical clustering methods and is based on the work of Hastie et al. (2009), which can be consulted for more information. Let $\boldsymbol{x}_j = (x_{j1}, \ldots, x_{jp})$ denote the observed values of p attributes (i.e., variables) for objects $j = 1, \ldots, n$. Let $d_l(x_{il}, x_{jl})$ denote the dissimilarity between the l-th attribute observed for objects i and j. For example, a commonly used dissimilarity measure for quantitative attributes is the squared distance $d_l(x_{il}, x_{jl}) = (x_{il} - x_{jl})^2$. For the non-hierarchical clustering methods used in this thesis, the overall dissimilarity between

the observed attributes of objects i and j can be written as

$$d(\boldsymbol{x}_i, \boldsymbol{x}_j) = g\left(\sum_{l=1}^{p} d_l(x_{il}, x_{jl})\right), \tag{2.15}$$

where g is some function applied to the sum of the dissimilarity measures over the observed attributes. Two overall dissimilarity measures are considered in this thesis: the first is the Euclidean distance, for which the distance measures are squared distances and the function g is the square root; the second measure is the squared Euclidean distance, for which the function g is the identity function. For the Euclidean and squared Euclidean distances, the attribute-specific squared differences in the sum of Equation (2.15) are assigned weights of one (i.e., each squared difference has the same weight). However, note that it is generally possible to assign different weights to the dissimilarity measures $d_l(\cdot, \cdot)$ (see Hastie et al., 2009 for further details).

In the clustering methods considered in this thesis, every cluster contains at least one object, and every object belongs to exactly one cluster. Assume the goal is to split the objects into k clusters, for some $k \in \mathbb{N}_{>0}$. Let $C(j)$ denote an *encoder* such that $C(j) = c$ denotes that object j is assigned to cluster c. The goal is to find cluster assignments that minimise some loss function. Generally, define the *within-cluster* point scatter $W(C)$ and *between-cluster* point scatter $B(C)$ as

$$W(C) = \frac{1}{2}\sum_{c=1}^{k}\sum_{C(j)=c}\sum_{C(i)=c} d(\boldsymbol{x}_j, \boldsymbol{x}_i),$$
$$B(C) = \frac{1}{2}\sum_{c=1}^{k}\sum_{C(j)=c}\sum_{C(i)\neq c} d(\boldsymbol{x}_j, \boldsymbol{x}_i),$$

such that the *total* point scatter T (which is independent of the cluster assignments) can be computed as

$$\begin{aligned} T &= \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} d(\boldsymbol{x}_j, \boldsymbol{x}_i) \\ &= W(C) + B(C). \end{aligned}$$

Partitioning methods aim to minimise the within-cluster point scatter, which, rewrit-

ten as $W(C) = T - B(C)$, is the same as maximising the between-cluster point scatter. Note that these goals reflect the intuitive objective of assigning similar objects to the same cluster and more dissimilar objects to different clusters.

2.8.1.1 k-means

The k-means algorithm is applied to objects with quantitative attributes to find k clusters to minimise the within-cluster point scatter when the squared Euclidean distance is used as the overall dissimilarity measure between objects. The standard algorithm to estimate this was proposed by Stuart Lloyd in 1957, and the method was published in Lloyd (1982). The `kmeans` function in `R` uses the algorithm provided by Hartigan and Wong (1979), which is based on some preliminaries from Hartigan (1975).

The k-means algorithm computes the squared Euclidean distance between attribute values $\boldsymbol{x}_j = (x_{j1}, \ldots, x_{jp})$ observed for objects $j = 1, \ldots, n$, as

$$d(\boldsymbol{x}_j, \boldsymbol{x}_i) = \sum_{l=1}^{p} (x_{jl} - x_{il})^2 = ||\boldsymbol{x}_j - \boldsymbol{x}_i||^2.$$

The within-cluster point scatter is computed as the *within-cluster sum of squares*

$$\begin{aligned} W(C) &= \frac{1}{2} \sum_{c=1}^{k} \sum_{C(j)=c} \sum_{C(i)=c} ||\boldsymbol{x}_j - \boldsymbol{x}_i||^2 \\ &= \sum_{c=1}^{k} N_c \sum_{C(j)=c} ||\boldsymbol{x}_j - \bar{\boldsymbol{x}}_c||^2, \end{aligned}$$

where $\bar{\boldsymbol{x}}_c = (\bar{x}_{c1}, \ldots, \bar{x}_{cp})$ is the mean vector associated with the c-th cluster, and $N_c = \sum_{j=1}^{n} \mathbb{1}(C(j) = c)$ is the number of objects assigned to cluster c, where $\mathbb{1}$ denotes an indicator function such that $\mathbb{1}(C(j) = c) = 1$ if $C(j) = c$ and $\mathbb{1}(C(j) = c) = 0$, otherwise. Thus, the entries of the mean vector for each cluster are computed as $\bar{x}_{cl} = \frac{1}{N_c} \sum_{j \in c} x_{jl}$, for $c = 1, \ldots, k$ and $l = 1, \ldots, p$.

The k-means algorithm aims to minimise the within-cluster sum of squares; i.e., the goal is to find the clustering C^* such that $C^* = \min_{C} W(C)$. For cluster centres

$(\boldsymbol{m}_1, \ldots, \boldsymbol{m}_k)$, the algorithm solves the optimisation problem

$$C(j) = \underset{1 \leq c \leq k}{\operatorname{argmin}} ||\boldsymbol{x}_j - \boldsymbol{m}_c||^2. \tag{2.16}$$

The k-means algorithm is initialised with k randomly chosen objects $\left(\boldsymbol{m}_1^{(0)}, \ldots, \boldsymbol{m}_k^{(0)}\right)$ to form the cluster centres in the optimisation problem from Equation (2.16), which is solved to obtain the initial clustering $C^{(0)}$. In the t-th iteration, for $t = 1, 2, \ldots$, the algorithm is carried out in the following steps:

1. For the current cluster assignment $C^{(t-1)}$ consisting of k clusters, compute the cluster means $\bar{\boldsymbol{x}}_c^{(t-1)}$ and make them the cluster centres, i.e., let $\boldsymbol{m}_c^{(t-1)} = \bar{\boldsymbol{x}}_c^{(t-1)}$ for $c = 1, \ldots, k$.

2. For the current cluster means $\left\{\boldsymbol{m}_1^{(t-1)}, \ldots, \boldsymbol{m}_k^{(t-1)}\right\}$, find the new clustering $C^{(t)}$ that solves the optimisation problem in Equation (2.16) by assigning each object to its closest cluster mean.

3. Increase t by one and start over with step 1 until the assignments no longer change.

With each iteration, the within-cluster sum of squares is reduced (e.g., see Hastie et al., 2009). However, this only guarantees that a local optimum is found; it does not guarantee that a global optimum will be found. Hence, it is common practice to run the algorithm multiple times with different randomly chosen initial clusterings and choose the result with the smallest within-cluster sum of squares to increase the chance of finding a global minimum.

Lastly, note that using the squared values in the optimisation problem in Equation (2.16) implies that large distances have a stronger influence and the method lacks robustness against outliers (Hastie et al., 2009). Hence, when one suspects outliers to be present, a more robust method should be used. One option is the k-medoids algorithm, which is presented next.

2.8.1.2 k-medoids (partitioning around medoids)

Kaufman and Rousseeuw (1987) propose the partitioning around medoids (PAM)

algorithm (also known as the k-medoids algorithm) and accredit the initial idea to Vinod (1969). The goal is to form k clusters around representative objects, called medoids, for which the average dissimilarity to other objects in the cluster is minimal. For a set of medoids $(\boldsymbol{m}_1, \ldots, \boldsymbol{m}_k)$, the clusters are formed by solving the optimisation problem

$$C(j) = \underset{1 \leq c \leq k}{\operatorname{argmin}}\, d(\boldsymbol{x}_j, \boldsymbol{m}_c). \tag{2.17}$$

Note that the k-medoids optimisation problem from Equation (2.17) is more general than that of the k-means algorithm, as the k-medoids algorithm can consider any overall dissimilarity, while the k-means algorithm uses the squared Euclidean distance.

The k-medoids algorithm is initialised with k randomly chosen objects $\left(\boldsymbol{m}_1^{(0)}, \ldots, \boldsymbol{m}_k^{(0)}\right)$ to form the initial cluster centres. The within-cluster point scatter is minimised by assigning each object to its closest cluster centre; i.e., the clustering $C^{(0)}$ is obtained by solving the optimisation problem from Equation (2.17) for the initial cluster centres.

In the t-th iteration, for $t = 1, 2, \ldots$, the algorithm is carried out in the following steps:

1. For the current clustering $C^{(t-1)}$ consisting of k clusters, find the object in each cluster whose overall distance to the other objects in that cluster is minimal (i.e., the object for which the average distance or equivalently, the sum of the distances, to the other objects in the cluster is minimal). Thus, the representative object of cluster c is found as

$$j_c^* = \underset{j: C^{(t-1)}(j) = c}{\operatorname{argmin}} \sum_{C^{(t-1)}(i) = c} d(\boldsymbol{x}_j, \boldsymbol{x}_i),$$

   where $d(\cdot, \cdot)$ is an arbitrary distance function. Set $\boldsymbol{m}_c^{(t)} = \boldsymbol{x}_{j_c^*}$ as the new cluster centre, for $c = 1, \ldots, k$.

2. For the current cluster centres $\left(\boldsymbol{m}_1^{(t)}, \ldots, \boldsymbol{m}_k^{(t)}\right)$, find the new clustering $C^{(t)}$ that solves the optimisation problem from Equation (2.17).

3. Increase t by one and start over with step 1 until the assignments no longer change.

The PAM algorithm is implemented in `R` and available in the package `cluster` (Maechler et al., 2023) in the function `pam()`, where the distance function from steps 1 and 2 of the k-medoids algorithm can be specified as the Euclidean or Manhattan distances. In Chapter 5 the k-medoids algorithm is applied to univariate data, for which the Euclidean and Manhattan distances are equivalent.

2.8.1.3 Choosing the number of clusters

Note that non-hierarchical clustering methods such as the k-means or k-medoids algorithms require the specification of the desired number of clusters k. However, typically, the number of clusters that should be created for the specific problem is not obvious a priori. Hence, Rousseeuw (1987) proposes a method to find the best partition amongst the resulting clusterings obtained for different values of k. The idea is to create silhouettes for each object j when the data are split into k clusters. Silhouettes are statistics that compare the distances between an object j and the other objects in the same cluster with the distances between object j and the objects in the nearest neighbouring cluster.

Assume that the clustering method results in a clustering C. Let $\bar{d}(j, c)$ denote the average dissimilarity between object j and the objects in cluster c; i.e., $\bar{d}(j, c) = \frac{1}{N_c} \sum_{\{i:C(i)=c\}} d(\boldsymbol{x}_j, \boldsymbol{x}_i)$. Now, let $a(j)$ denote the average distance or dissimilarity between object j and all other objects in its cluster $C(j)$; i.e., $a(j) = \bar{d}(j, C(j))$. Let $B(j)$ denote the nearest neighbouring cluster of object j, i.e., the cluster for which the average distance to object j is $b(j) = \min_{c \neq C(j)} \bar{d}(j, c)$. The silhouette of object j is then computed as

$$s(j) = \frac{b(j) - a(j)}{\max\{a(j), b(j)\}},$$

which allows for the evaluation of the assignment of object j to cluster $C(j)$. Note that $s(j) = 0$ for $a(j) = b(j)$. If the average distance between object j and all objects in cluster $C(j)$ is similar to that between object j and all objects in cluster $B(j)$,

its silhouette $s(j)$ will be close to zero. On the other hand, if the average distance between object j and the objects in cluster $B(j)$ is much larger than the average distance between object j and the other objects in cluster $C(j)$, then the silhouette $s(j)$ will be close to one. Similarly, negative values of $s(j)$ indicate that object j is, on average, closer to the objects in cluster $B(j)$ than to objects in cluster $C(j)$, suggesting that object j might have been falsely assigned to cluster $C(j)$.

Lastly, the average silhouette width for all k clusters can be computed as

$$\bar{s}_k = \frac{1}{n} \sum_{j=1}^{n} s(j).$$

Using this method, one can argue that the optimal number of clusters k is the one that results in the largest average silhouette width $\bar{s}_k$.

Another option for checking the optimal number of clusters (which is less formal, but commonly applied in combination with k-means) is the *elbow* method, which is attributed to Thorndike (1953) and based on a plot of the within-cluster sum of squares values for clustering assignments with different numbers of clusters k. Note that the within-cluster sum of squares generally decreases as the number of clusters k increases (e.g., see Hastie et al., 2009). When clustering is meaningful, splitting the dataset into two clusters should lead to a substantial reduction of the within-cluster sum of squares. Similarly, having a few meaningful clusters might lead to further substantial reductions. However, once a sufficiently large number of clusters has been selected, one would anticipate the reductions in the within-cluster sum of squares to be small. The change from a substantial decrease to a smaller decrease as one increases the number of clusters looks like an elbow in the within-cluster sum of squares plot. Hence, the plot can be used to pick the optimal number of clusters by identifying the location of the elbow.

As an example for using the average silhouette widths and the elbow method, consider the famous *iris* dataset collected by Anderson (1936). The dataset contains 50 measurements obtained for each of three species of the iris flower: Setosa, Versicolor, and Virginica. Figure 2.1 shows exploratory plots for the variable *petal width*. The plot on the left shows a histogram over all observations, which one could consider if it

were unknown that the observations are from three different species. The plot on the right shows the distribution of the observed petal widths by species and reveals that the species Setosa tends to have the smallest petal widths, followed by Versicolor, and Virginica tends to have the largest petal widths. However, it should be noted that the differences in petal width between the species Versicolor and Virginica tend to be smaller than between Setosa and Versicolor, with a range of petal widths being observed for both Versicolor and Virginica.

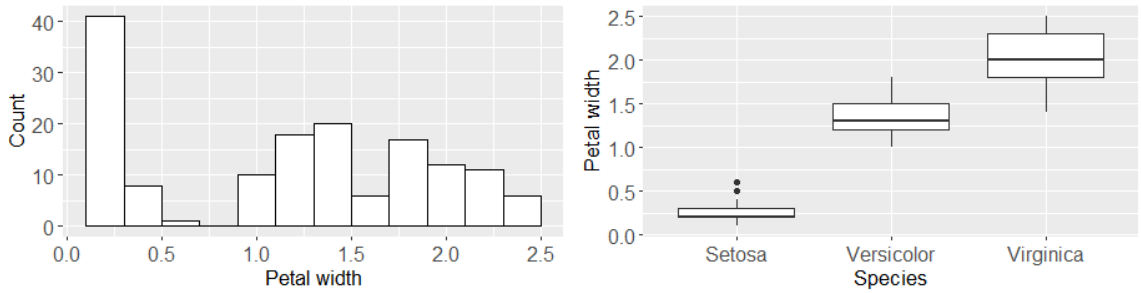


Figure 2.1: Iris dataset: A histogram of the petal width over all observations (left) and boxplots showing the distribution of the observed petal widths by species (right).

Figure 2.2 shows the results from k-means clustering applied to the variable petal width for different numbers of clusters k, pretending that the different species were unknown. The plot on the left shows that the within-cluster sum of squares reduces substantially when moving from one to two clusters and from two to three clusters. Increasing the number of clusters further leads to a much smaller decrease in the within-cluster sum of squares values, which is visible as an "elbow" in the plot at $k = 3$. The plot on the right shows that the average silhouette width is largest for $k = 3$, although it is only marginally larger than for $k = 2$. When splitting the observations in the iris dataset into two clusters according to the variable petal width, the 50 observations of the species Setosa are in one cluster, and the 100 observations from the species Versicolor and Virginica are in the other cluster. Splitting the dataset into three clusters according to the variable petal width, the 50 observations of the species Setosa again form one cluster, 48 observations of the species Versicolor and four observations of the species Virginica form a second cluster, and the third cluster is formed of the remaining 46 observations of the species Virginica and two observations of the species Versicolor. Hence, in this case, the clustering is meaningful and the small difference in the average silhouette widths for $k = 2$ and $k = 3$ is explained by

the observed petal widths of the species Versicolor and Virginica being more similar to one another (see Figure 2.1).

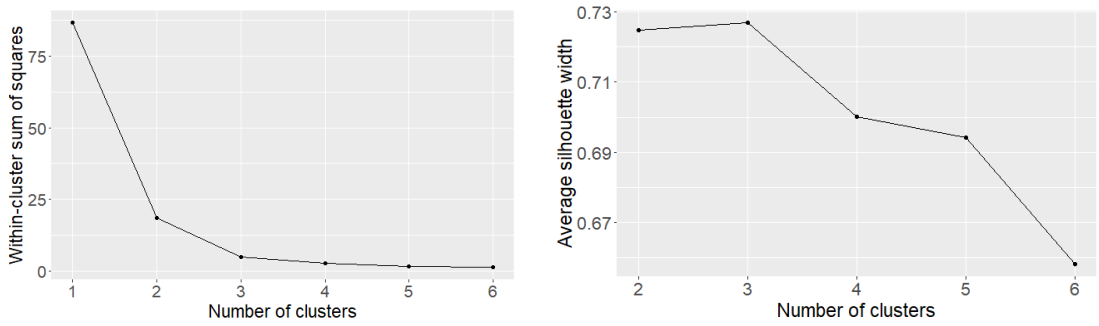


Figure 2.2: Iris dataset: Within-cluster sum of squares (left) and average silhouette width (right) for the clusterings obtained by applying the k-means algorithm to the observations of the variable petal width.

The average silhouette widths should be used as a less subjective method for selecting the number of clusters. However, the average silhouette width can only be computed when the data are split into at least two clusters. Hence, the method cannot be used to check whether the data should be split into clusters at all. Instead, the within-cluster sum of squares plots can be used to check if the within-cluster sum of squares reduces substantially when splitting the data into two clusters, compared to when all objects are assigned to one cluster (e.g., see Figure 2.2). Another example plot for this scenario can be found in Figure 3.7 of Chapter 3, where the average silhouette widths are largest for two clusters and the within-cluster sum of squares plots show that the within-cluster sum of squares reduce substantially when moving from one to two clusters.

2.8.1.4 Rand index

Rand (1971) proposes an index to compare how similar two clusterings are, and the index is commonly referred to as Rand index. Assume the objective is to split n objects into clusters. Consider two clusterings $C = \{c_1, \ldots, c_k\}$ and $C' = \{c'_1, \ldots, c'_{k'}\}$, where k and k' denote the respective number of clusters. The Rand index is computed as

$$r(C, C') = \left(\sum_{i<j}^{n} \gamma_{ij} \right) \Big/ \binom{n}{2},$$

where $\gamma_{ij} = 1$ if for objects i and j, either

- $C(i) = C(j)$ and $C'(i) = C'(j)$ (the objects are assigned to the same cluster in both clusterings), or
- $C(i) \neq C(j)$ and $C'(i) \neq C'(j)$ (the objects are assigned to different clusters in both clusterings),

and $\gamma_{ij} = 0$, otherwise.

The Rand index reflects the comment of Rand (1971) that "clusters are defined just as much by those points which they do not contain as by those points which they do contain", as the similarity measure γ_{ij} not only takes on a value of one when the two objects i and j are in the same cluster for both clusterings C and C', but also when the two objects are in different clusters for both clusterings. The similarity measure γ_{ij} is computed for a total of $\binom{n}{2}$ pairs (i, j) with $i < j$, and the Rand index is the proportion of pairs for which $\gamma_{ij} = 1$. Two clusterings C and C' are identical if $\gamma_{ij} = 1$ for all pairs (i, j), in which case $\sum_{i=1}^{n} \gamma_{ij} = \binom{n}{2}$ so that the Rand index takes on a value of one, i.e., $r(C, C') = 1$. The only possible scenario for $r(C, C') = 0$ is where one clustering assigns all objects to one cluster, while the other consists of n clusters each containing only one object. For other non-identical clusterings, the Rand index will be between zero and one and is closer to one the more similar the two clusterings are.

2.8.2 Single linkage agglomerative clustering

Hierarchical clustering methods do not lead to a single assignment of objects to clusters, but instead provide a hierarchy of clusters that can be explored visually. There are two types of hierarchical clustering: agglomerative and divisive clustering (e.g., see Giordani et al., 2020). These are recursive methods, where objects are partitioned into clusters in a hierarchical fashion where the former follows a bottom-up approach (starting with single-object clusters that are merged) and the latter a top-down approach (starting with all objects in one cluster that is divided into sub-clusters).

Agglomerative clustering algorithms start with n singleton clusters, i.e., with one cluster for each object. In each of $(n-1)$ steps, the two clusters with the smallest dissimilarity (according to some dissimilarity function) are merged. In the $(n-1)$-th step, the two remaining clusters are merged so that all observations are in one cluster. The clusterings in this hierarchy can be presented in a binary tree, where the nodes present the clusters at each level. The root node (at the top of the tree) represents the entire dataset (i.e., all objects are in one cluster), and the n terminal nodes (at the bottom of the tree) are the singleton clusters from the start of the agglomerative method. Each non-terminal node has two daughter nodes, as the nodes higher up in the tree are formed by merging two clusters from the level below. The tree can be plotted so that the height of each node is proportional to the dissimilarity between its two daughter nodes, where the terminal nodes are plotted at zero height. The resulting tree plot is called a *dendrogram* and can be used to decide which hierarchy level should be used to obtain the final clustering. For further details, see, for example, Hastie et al. (2009).

This thesis uses single linkage agglomerative clustering. Again, consider n objects for which p variables are observed, so that $\boldsymbol{x}_j = (x_{j1}, \ldots, x_{jp})$ denotes the attributes observed for the j-th object, for $j = 1, \ldots, n$. Single linkage (SL) agglomerative clustering computes the distance between two clusters as the smallest distance between any two objects from each cluster; i.e., the dissimilarity between two clusters c_1 and c_2 is computed as

$$d_{\text{SL}}(c_1, c_2) = \min_{i \in c_1, j \in c_2} d(\boldsymbol{x}_i, \boldsymbol{x}_j), \tag{2.18}$$

where $d(\cdot, \cdot)$ is some dissimilarity measure and the clusters are merged following the agglomerative approach outlined above.

A benefit of single linkage agglomerative clustering is that it can be used to identify potential singleton outliers (Everitt et al., 2011). Since the singleton outliers will be far from their nearest neighbours, they can be identified as the terminal nodes with the greatest heights. Chapter 3 applies single linkage clustering to check for potential outliers before applying k-means clustering.

Chapter 3

National lockdowns in England: The same restrictions for all, but do the impacts on COVID-19 mortality risks vary geographically?

3.1 Introduction

The COVID-19 pandemic is the deadliest respiratory disease pandemic since the "Spanish" influenza in 1918, and it is one of at least three detected coronaviruses that have spread since the year 2000 (Morens et al., 2020). The disease is zoonotic (transmittable between humans and animals), and it is believed to have originated from the Hunan seafood market (Shereen et al., 2020) in Wuhan, China. A previous outbreak of Severe Acute Respiratory Syndrome (SARS) coronavirus in 2002 also had its first recorded cases in China, in the Guangdong Province (Monagin et al., 2018) which, just as Wuhan, falls into a region that Jones et al. (2008) have identified as a high relative risk area for human emerging infectious diseases caused by zoonotic pathogens from wildlife. Taylor et al. (2001) have shown that a majority of infectious

organisms fall into the zoonotic category, and Karesh et al. (2005) and Swift et al. (2007) have indicated a link between wildlife markets and emerging infectious diseases, calling for regulations on the trade of wildlife. Yet, without such regulations or an effective early response strategy in place, the newly formed coronavirus was able to grow from local chains of infection to a worldwide pandemic, as declared by the World Health Organisation (WHO) on 11th March 2020 (Director-General, 2020).

As of 24th March 2022, there have been over 6.1 million recorded deaths worldwide that were linked to COVID-19, from over 475 million recorded cases of COVID-19 infections (https://coronavirus.jhu.edu/map.html). While deaths and severe disease progressions represent the most immediate and dire consequences of the spread of the virus, the pandemic has also had a strong environmental and socio-economic impact (Bashir et al., 2020). Furthermore, while it is still too early to fully predict the lasting impact of the pandemic, current research suggests possibilities of long-term effects on physical health (Sudre et al., 2021), as well as mental health (Usher et al., 2020). It is also likely that the economic decline and possible lack of social contacts during the pandemic will cause additional hardship on the mental well-being of individuals.

Some key scientific challenges related to the virus have been to model and predict the spread of infections (e.g., Dong et al., 2020, Lee et al., 2022), to identify the factors that are associated with a higher risk of displaying severe symptoms (e.g., Rashedi et al., 2020, Williamson et al., 2020, Wolff et al., 2021), and to identify impacts on healthcare (e.g., Remuzzi and Remuzzi, 2020). These findings were particularly impactful during the early stages of the pandemic, as they provided governments with the scientific knowledge necessary for developing strategies to contain the virus to assist the stressed healthcare sector by reducing the number of patients with severe disease progression. Some of the earliest strategies included increased opportunities to maintain good hygiene through hand sanitisation in public spaces, enforcing social distancing rules and face masks in enclosed settings. Additionally, those who tested positive for an infection or had recently been in close contact with someone who tested positive were asked to isolate, and people travelling from regions with higher infection rates were sometimes asked to obey quarantine rules upon arrival. Having been

identified as particularly vulnerable to displaying severe symptoms, older adults and those with pre-existing medical conditions received additional protection by regular testing of the care home and medical staff. The more vulnerable groups were also given priority in getting their vaccinations once available.

Most governments implemented restrictions to constrain the spread of the virus, especially when rising numbers of severely affected individuals who needed intensive care started to overwhelm the national healthcare sectors. A common intervention was to implement regional or national lockdowns. While the exact restrictions during lockdown differ by country, in general, a lockdown is "a temporary condition imposed by governmental authorities (as during the outbreak of an epidemic disease) in which people are required to stay in their homes and refrain from or limit activities outside the home involving public contact" (Merriam-Webster Online Dictionary, 2022).

In England, the region of this study, three national lockdowns occurred between the start of 2020 and October 2021. The lockdowns ranged from 26th March to 12th May 2020 (48 days), from 5th November to 2nd December 2020 (29 days), and from 5th January to 28th March 2021 (83 days). Lockdowns are implemented to break the chains of infection, but they come with a large social and economic burden for society. The global economy experienced a collapse during the pandemic, which might have been partly linked to governmental restrictions and voluntary consumer decisions (Goolsbee and Syverson, 2021). The financial crisis posed yet another possible threat to the mental well-being of many individuals, adding to the burden of limited social contacts and fear of infection. While the imposed lockdowns did not appear to have "uniformly detrimental effects on mental health" (Prati and Mancini, 2021), some studies report an increase in depressive symptoms and worsened mental well-being amongst adolescents (Thorisdottir et al., 2021), who experienced a particularly strong impact of the pandemic, as schools were closed for extended periods.

The overriding goal of this study is to investigate spatio-temporal trends in COVID-19 mortality risks following the implementation of three national lockdowns in England to identify geographical differences in the impact of lockdown. Before introducing the methodology used to conduct the analysis, other studies that have previously investigated temporal trends in COVID-19 death counts or mortality risks

following lockdowns are reviewed.

Palladino et al. (2020) and Palladino et al. (2021) fit interrupted time series (ITS) models to the number of deaths between late February and early May of 2020 in Italy and between late January and mid-August of 2020 in Italy, France, Spain, and the UK, respectively. The studies showed a decreasing trend in the number of deaths upon implementing a national lockdown. However, they did not give any insights into spatial patterns within the countries, nor did the latter study allow a comparison between the countries, as the response variable was raw counts of deaths rather than standardised mortality rates. Conyon et al. (2020) used a Difference-in-Differences model from econometrics to show that the number of deaths per million was significantly lower in Denmark and Norway, where a "hard lockdown" policy was put in place, compared to Sweden, where no such "hard lockdown" was implemented. Gerli et al. (2020) estimated the raw numbers of deaths in 29 European countries by fitting a model that consisted of a cubic polynomial and parametric growth curve, using the observed mortality distribution within the first 17 days of the outbreak and the date of the implementation of national lockdown as predictor variables. Their findings suggested a correlation between earlier lockdowns and lower numbers of deaths. However, it should be pointed out that lockdown measures differed by country, as the lowest level of "lockdown" denoted a travel ban only, which makes an equal comparison of the lockdowns in different countries a questionable procedure.

Coccia (2021) compared data for countries with a shorter period of first lockdown (Austria, Portugal, Sweden) to countries with a longer period of first lockdown (France, Italy, Spain). The study claimed that "a longer period of national lockdown to constrain the diffusion of COVID-19 does not seem to be associated with a significant reduction of infected cases on population and of fatality rates in society". However, there are some key issues inherent in the analysis. Firstly, the fatality rate (in %) of COVID-19 was measured as deaths divided by the total number of infected individuals in each country, which reduces the fatality rates for countries with higher testing capacities, as more people with less severe symptoms could be tested, therefore generating more positive test results on patients with less severe disease progression. Secondly, the statistical tests in the study are questionable, as a t-test was performed

to compare the mean fatality rates from April to August 2020 between the countries without considering the initial fatality rates (before the lockdown). Recall that Spain and Italy were the European countries most severely affected during the early months of the pandemic, which will likely impact these results.

Other studies have analysed the effect of lockdowns at regional levels. Silva et al. (2020) have fit an ITS model to data from four state capitals in Brazil, each of which had implemented regional lockdowns over some period in April to June 2020, while most parts of the country did not implement strict regulations during that time frame. Specifically, their study is based on ordinary least squares ITS models, with log-transformed daily numbers of deaths as the response variable. While an observed change in the slope of their fitted model suggests a decreasing trend in the daily number of deaths upon implementing the lockdown, the results do not provide any insight into the underlying spatial patterns of deaths. Further, these trends could not be compared for the different cities, as raw counts were used as the response variable rather than standardised rates of deaths.

dos Santos Siqueira et al. (2020) used the *Joinpoint Regression Program* to estimate daily mortality rates (per 100,000 people) in the autonomous communities of Spain between 14 March and 25 April 2020, which constitutes the time frame of Spain's first national lockdown in the pandemic. The methodology is similar to that of ITS, the only difference being that Joinpoint Regression identifies the time points based on changes in the data. In contrast, in an ITS model, the researcher specifies the time points of interest. On average, the observed points of change suggested that the mortality rates started to decrease 18.33 days after the lockdown was introduced, and estimates for different autonomous communities are compared in the paper. The authors state that the models were fitted with autonomous communities set as strata, suggesting independent estimates for the different communities. Considering the underlying nature of COVID-19 as an infectious disease, spatial autocorrelation amongst the autonomous communities would appear a more appropriate assumption since infection rates would likely be more similar for neighbouring autonomous communities than for autonomous communities that are further apart.

Note that the studies by Silva et al. (2020), Palladino et al. (2020), dos San-

tos Siqueira et al. (2020), and Coccia (2021) additionally consider the number of cases of infection as a response variable, the results of which are ignored in this review as their reliability could be questionable, due to limited testing capacities, especially during the first months of the pandemic. That is why this English study focuses on death counts.

Davies et al. (2020) fit Susceptible-Exposed-Infected-Removed (SEIR) models to COVID-19 data from the UK. They inspect the potential impact of combinations of different control measures (school closures, physical distancing, shielding of people aged 70 years and older, self-isolation of symptomatic cases) on the number of deaths due to COVID-19. They construct an age-specific matrix that captures UK social mixing patterns amongst age groups to simulate the susceptible, exposed, infected, and removed status for individuals in the population. Their simulations suggest that an unmitigated COVID-19 pandemic would result in 350,000 deaths in the UK up to December 31st, 2021, while lockdowns implemented after 1,000 intensive care unit beds were filled nationally could reduce this number to 50,000 deaths. It should be noted that the prediction intervals for the number of deaths are fairly wide. Yet, the results allow to compare different scenarios for combinations of control measures, suggesting that lockdowns would be most effective in reducing the number of deaths. Davies et al. (2021) fit the same model to compare a baseline scenario of no tiered restrictions or lockdown interventions to implementing tiers and to lockdowns similar to the ones that had been introduced in Northern Ireland (non-essential retail remained open, household bubbles of up to ten people from two households) and Wales (non-essential retail was closed, no mixing between households). Their findings indicate differences in the number of deaths by region in England. Further, they suggest that the types of lockdown introduced in Northern Ireland and Wales would be more effective in reducing the cumulative number of deaths than the tier system implemented in England. Sartorius et al. (2021) fit an SEIR model, which considers population mobility data from England. They quantify the impact of key socio-demographic risk factors on COVID-19-related mortality risk, showing that mortality risk was elevated for areas with higher proportions of the population aged 70 years and older and for areas with higher proportions of older adults living in deprivation. Overall, their

findings suggest differences in mortality risk at a small area scale in England, which should be investigated further.

This study aims to explore the temporal trends in COVID-19 mortality risks after the introduction of the national lockdowns, using weekly counts of mortalities on a medium area scale of local authority districts (LADs) in England. A common exploratory measure of disease risk is the *standardised mortality ratio* (SMR, recall Section 2.1), computed as the observed counts of deaths divided by the expected counts. These expected counts adjust for LADs having different population sizes and demographics, such as age and sex. Further details are provided in Section 3.3. While SMRs can provide a first impression of mortality risk, they should be viewed critically as they are prone to substantial variability. For this study's data, this variability is explained by the relatively low weekly observed and expected counts of COVID-19 deaths at a local authority level. In general, spatio-temporal models can reduce some of the random variability in the data, as they capture spatio-temporal trends and autocorrelation and borrow strength in risk estimates between neighbouring data points. In the context of epidemiology, spatial autocorrelation might be induced by unmeasured confounding (some variable that is not included as a predictor in the model explains part of the variability in the data), neighbourhood effects (similar rates of disease in areas of proximity), or grouping effects (when people of similar demographics live close to each other). Similarly, temporal autocorrelation could be caused by the fact that largely the same population are at risk in consecutive periods. For infectious diseases, spatial and temporal autocorrelation is even more apparent as the disease is passed on between people interacting in space and time. The specific model applied in this study is a Poisson log-linear model developed to estimate the long-term effects of air pollution on respiratory hospital admissions in Greater London (Rushworth et al., 2014). Section 2.7.4 provides a general description of the model, and its application to the English COVID-19 mortality data will be discussed in Section 3.3. The model captures the spatio-temporal trends and correlations in the data, resulting in smoothed and, therefore, more reliable estimates of relative mortality risk for each spatial and temporal unit while accounting for the underlying age and sex demographics. From these estimated mortality risks, scaled risks can be

computed for each area, comparing the estimated risks during and after the lockdown to those from the last week preceding the lockdown. Hence, the scaled risk can be used to fairly compare how the impact of lockdown varies for different areas across England.

To my knowledge, this study is the first comprehensive investigation of spatio-temporal trends in COVID-19 mortality risks following the implementation of lockdowns in England, and it is the first to consider all three national lockdowns that have occurred thus far. The main questions that will be answered are the following:

1. How long after the implementation of lockdown did mortality risks reduce at a national level, and did this vary by lockdown?

2. How did the temporal trends in mortality risks differ by region in England?

3. Which local authorities were exposed to the highest average risks in the weeks after lockdown?

4. Which local authorities shared similar temporal trends in mortality risks?

The remainder of this chapter is organised as follows. Section 3.2 introduces the COVID-19 mortality data used in this study. Section 3.3 describes the methodology that is applied. Section 3.4 displays the results of the study, and Section 3.5 provides a discussion of the results, particularly how they could be used to aid the government in future decision-making when considering implementing a national or regional lockdown.

3.2 Motivating Study

The time frame of this study is from 1st February 2020 (the week of the first registered death due to COVID-19 in England) to 14th May 2021 (seven weeks after the third lockdown was lifted), and the study region is mainland England which is partitioned into local authority districts (LADs). The average population size for a single LAD is 179,945, with sizes ranging from 9,721 to 1,141,816 people. Since the LADs are not large, the geographical risk analysis also considers the nine regions of England that

LADs nest exactly within. The regions are East of England, East Midlands, London, North East, North West, South East, South West, West Midlands, and Yorkshire and The Humber. Figure 3.1 presents a map of the nine regions of England, which also shows the outlines of the LADs.

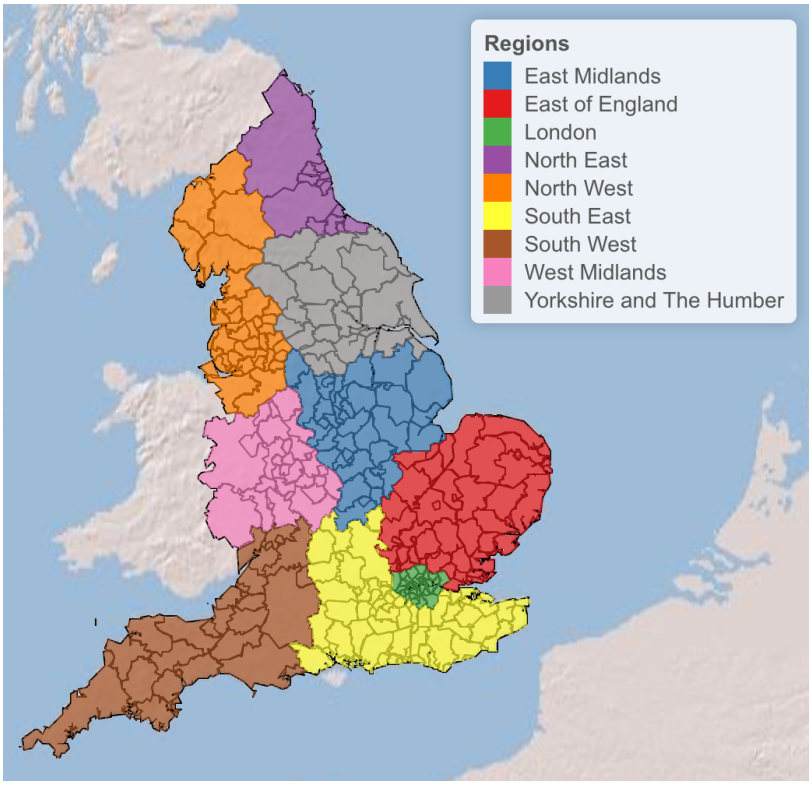


Figure 3.1: The nine regions of England on a map with outlines of the LADs.

The data contain the number of deaths due to COVID-19 for each of a total of 312 LADs and 67 weeks. The number of deaths in LAD j and week t is denoted Y_{jt}. During the observed time frame, there were three national lockdowns in England over the following ranges:

- Lockdown 1: 26/03/2020 - 12/05/2020 (48 days),
- Lockdown 2: 05/11/2020 - 02/12/2020 (28 days),
- Lockdown 3: 05/01/2021 - 28/03/2021 (83 days).

It should be noted that the days of the week when lockdowns were implemented do not align with the first day of the week in the mortality data, where each week is defined to range from Saturday to Friday. Therefore, I consider each week in the death data with at least four days of lockdown as a week during the lockdown, while

any week with three or fewer lockdown days falls outside of lockdown. This mismatch causes a slight shift in lockdowns for the data used in this study, as the lockdowns get assigned to 28/03/2020 - 15/05/2020, 07/11/2020 - 04/12/2020, and 02/01/2021 - 26/03/2021.

Deaths in England are registered on the General Register Office's Registration Online system (https://www.gov.uk/general-register-office). The Office of National Statistics (ONS, 2021b) has provided the occurrences and registrations of deaths as weekly accumulated counts for the local authority districts in England and the weeks described above. While deaths from all causes are captured, this study only uses the numbers of deaths for which COVID-19 was mentioned on the death certificate. Figure 3.2 shows a plot of the raw counts of COVID-19 deaths across the whole of England for each week in the time frame of this study, and the weeks of national lockdown are highlighted in beige. The waves of the pandemic are reflected in the accumulated counts of deaths, with peaks in April 2020, November 2020, and January 2021.

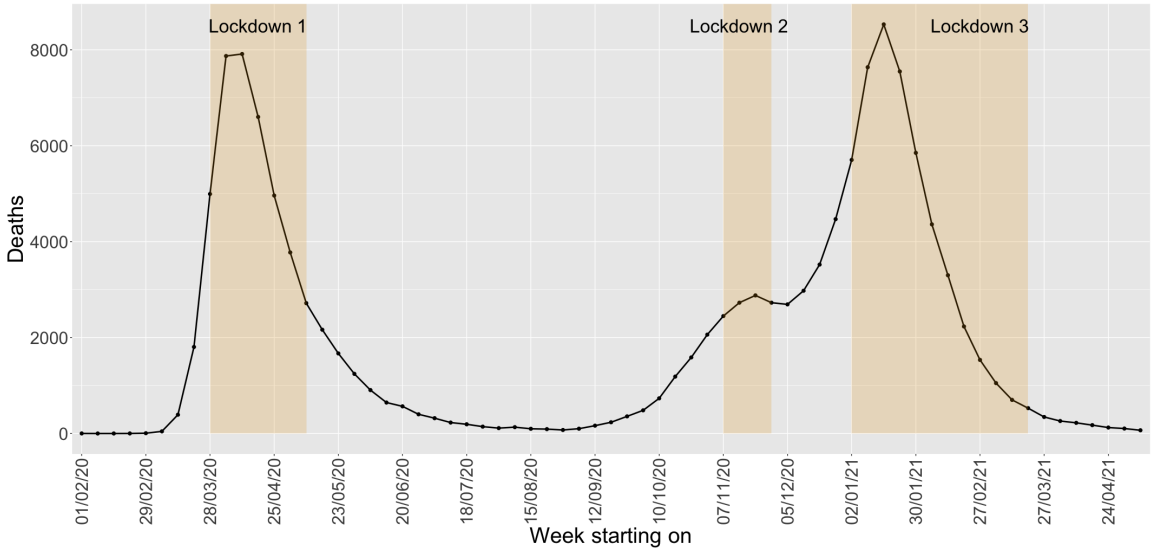


Figure 3.2: Counts of deaths across England. The weeks of national lockdown are highlighted in beige.

It should be noted that the data include only the deaths of those registered in England, and a person's death is counted towards the LAD in which they were registered, disregarding the actual place where the person died should these locations differ. Deaths with missing residences are not included. For this reason, the total number of deaths differs slightly from published figures. Further, this study uses

the dates of death occurrences rather than registrations of death, and differences between the dates of occurrences and registrations arise whenever there is a delay in the registration of deaths.

COVID-19 mortality risk differs by age and sex. For example, many more deaths occur amongst the older population than in younger age groups, and males tend to be at higher risk than females (Biswas et al., 2021). Hence, when comparing the COVID-19 mortality risk of different areas, the underlying age and sex demographics should be considered. In this study, standardised mortality ratios (SMRs, recall Section 2.1) are computed as initial risk estimates. First, the expected number of COVID-19 deaths is computed by applying indirect standardisation using age-sex-specific population data for mid-2020 (ONS, 2021c). Note that 2019 population data were used for seven LADs due to a lack of data for 2020. The expected numbers of deaths are computed from national age-sex-specific mortality rates for COVID-19, which were taken from https://coronavirus.data.gov.uk/details/deaths?areaType=nation&areaName=England on 30th August 2021. Let n_{ji} denote the number of people in LAD j from age-sex group i (e.g. females 0-4, females 5-9, etc.) from a total of I such groups, and ω_i denote the English national rate of COVID-19 mortality per 100,000 people for the i-th group. The expected weekly number of deaths $\tilde{e}_j$ is then computed as

$$\tilde{e}_j = \frac{1}{N} \frac{\sum_{i=1}^{I} n_{ji}\omega_i}{100,000}, \tag{3.1}$$

where N is the total number of weeks in the study, and $\tilde{e}_j$ does not change over time as time-varying data on population sizes or national mortality rates at a weekly scale are not available. Furthermore, this study's main objective is to investigate the temporal trends in mortality risks and hence, the expected counts are deliberately computed from the mortality rates over the entire study period so that the deviations in the weekly mortalities from the national levels over the entire period can be identified. However, the expected counts may not be on the same scale as the observed counts Y_{jt} (for example, national age-sex-specific mortality rates might not have been computed over the same period as the observed deaths in the data), and so the expected counts

are rescaled by

$$e_j = \tilde{e}_j \frac{\sum_{j,t} Y_{jt}}{\sum_{j,t} \tilde{e}_{jt}}. \tag{3.2}$$

The rescaling ensures that $\sum_{j,t} e_j = \sum_{j,t} Y_{jt}$, so that the observed and expected counts in total over all areas and weeks are the same. The standardised mortality ratio (SMR) is an exploratory measure of disease risk computed as

$$\text{SMR}_{jt} = \frac{Y_{jt}}{e_j}, \tag{3.3}$$

for LAD $j = 1, 2, ..., K(= 312)$ and week $t = 1, 2, ..., N(= 67)$. For example, $\text{SMR}_{jt} = 1.2$ suggests that area j has a 20% elevated risk in week t, compared to the national average risk over the study duration. Figure 3.3 presents the SMR by week and LAD with weeks of lockdown highlighted in beige, where the weekly average SMR over all LADs is shown by the black line. The dashed red line represents an SMR of one.

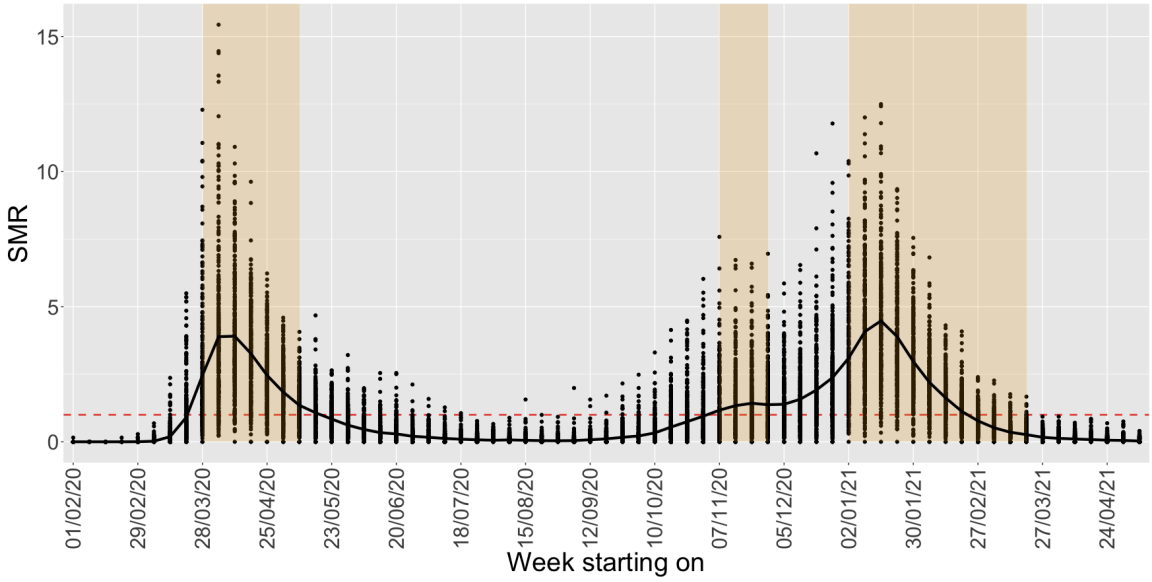


Figure 3.3: SMR by week for all LADs in England. The solid line shows the weekly average SMR over all LADs and the weeks of national lockdown are highlighted in beige.

The SMR values show two distinct waves, with peaks in April 2020 and January 2021. Note that this trend in time is similar to the pattern for the death counts observed in Figure 3.2. Overall, the SMR values appear to decline sometime after the introduction of lockdowns 1 and 3, and the national average SMR is below one (represented by the red dashed line) during many weeks in the summer and early

autumn of 2020 and late spring of 2021. The average SMR value across England does not appear to decrease much for the second lockdown, likely due to the lockdown being lifted after only four weeks.

Figure 3.4 displays the spatial pattern in the average SMRs for the LADs of England over the time frame of this study. The map suggests higher average SMRs for urban areas such as London, Birmingham, Manchester, and Liverpool. In contrast, the average SMRs appear lower for more rural areas such as England's South West, North West, and parts of the East of England. The average SMRs appear to change relatively smoothly across the map, suggesting spatial autocorrelation is likely in these data.

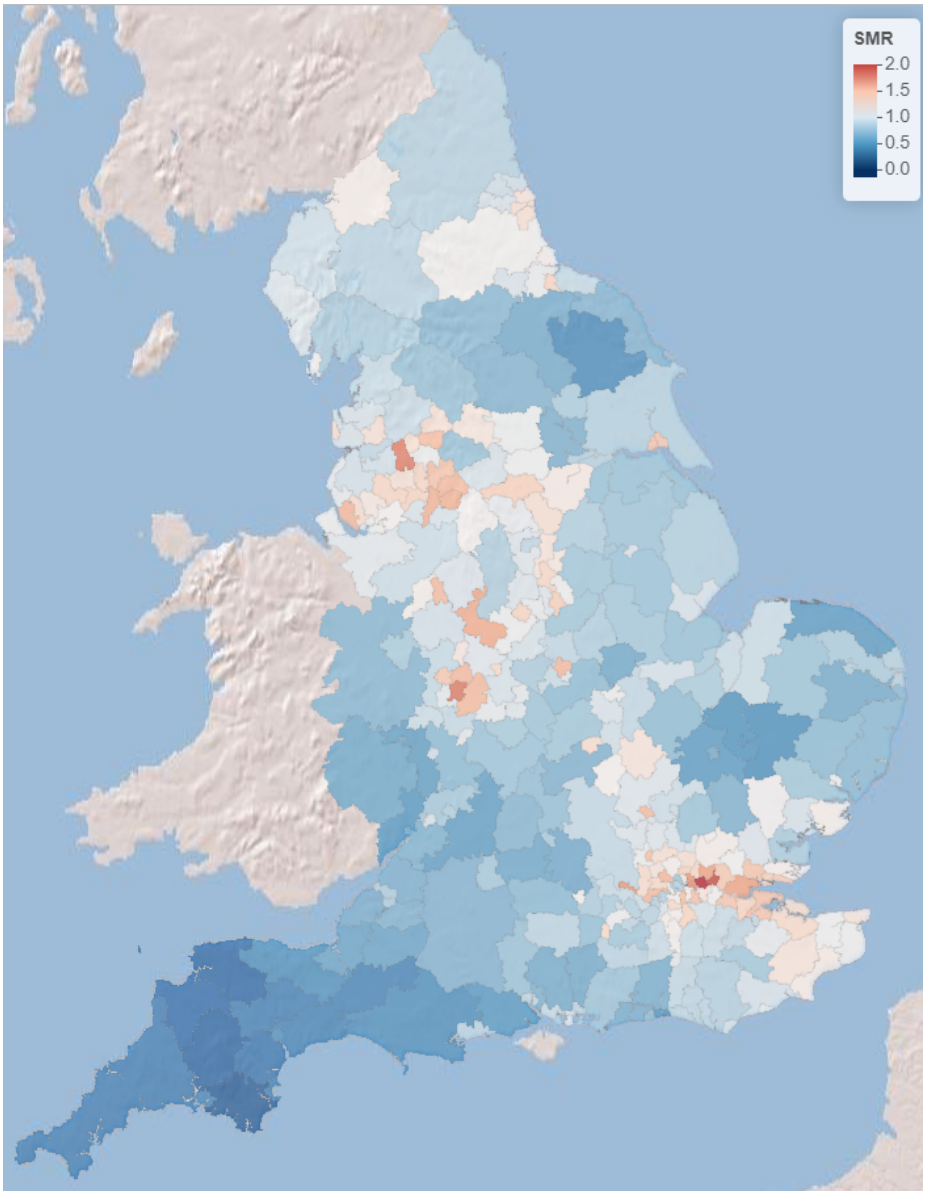


Figure 3.4: Average SMR over the entire study period, by LAD.

To verify the assumption of positive spatial autocorrelation a Moran's I test (Section 2.5.1.1) is performed for each week t in the study. The Moran's I test statistic for week t is computed as

$$I_t = \frac{K \sum_{j=1}^{K} \sum_{i=1}^{K} w_{ji}(\text{SMR}_{jt} - \overline{\text{SMR}_t})(\text{SMR}_{it} - \overline{\text{SMR}_t})}{(\sum_{j=1}^{K} \sum_{i=1}^{K} w_{ji}) \sum_{j=1}^{K} (\text{SMR}_{jt} - \overline{\text{SMR}_t})^2}, \tag{3.4}$$

where $\overline{\text{SMR}_t} = \frac{1}{K} \sum_{j=1}^{K} \text{SMR}_{jt}$ is the average SMR over all LADs in week t, and w_{ji} indicates whether or not LADs j and i share a border, i.e. $w_{ji} = 1$ if LADs j and i share a border, and $w_{ji} = 0$, otherwise. Note that the values $\{w_{ji}\}$ form the entries of a binary adjacency neighbourhood matrix $\boldsymbol{W}$ (recall Section 2.1), and the same neighbourhood structure is used by the statistical model described in Section 3.3. Moran's I test statistic cannot be computed for the weeks starting 08/02 and 15/02/2020, as no COVID-19 deaths had occurred during those weeks, causing the SMRs to be 0 for all LADs. For the other 65 weeks in the data, the Moran's I statistics take on values between -0.021 and 0.661, with a mean value of 0.257. Using Monte Carlo simulation, a permutation test is performed based on 10,000 random permutations of the data to test if the spatial autocorrelations in each week are significant. In each of the 10,000 iterations, the SMR values for the week in question are randomly assigned to the LADs, and Moran's I statistic is computed for each permuted set. This distribution generated under independence is then compared to the original statistic computed for the observed data from that week to yield a p-value against independence (recall Section 2.5.1.1).

A Bonferroni correction (Haynes, 2013, also see Shaffer, 1995) is applied to retain a significance level of 0.05 in the presence of multiple tests so that the significance threshold is $0.05/65 \approx 0.00077$. This results in rejecting the null hypothesis of no spatial autocorrelation for 48 of the 65 weeks (73.95%). Hence, it can be concluded that spatial autocorrelation is present in most of the weeks in the data, so a model that can account for spatial autocorrelation should be fitted. For additional explanations of the methods applied and instructions on performing the analysis, see Lee (2020) for a tutorial on spatio-temporal disease risk modelling in R.

Temporal autocorrelation in the SMR values is formally checked by computing

temporal autocorrelation coefficients for each LAD over the weeks in the study. Following the definition from Section 2.5.2, a temporal sample autocorrelation coefficient at lag l for LAD j in the data can be computed as

$$\hat{r}_{jl} = \sum_{t=1}^{N-l} \frac{(\text{SMR}_{jt} - \overline{\text{SMR}}_j)(\text{SMR}_{j(t+l)} - \overline{\text{SMR}}_j)}{\sum_{t=1}^{N}(\text{SMR}_{jt} - \overline{\text{SMR}}_j)^2}, \tag{3.5}$$

where $\overline{\text{SMR}}_j = \frac{1}{N}\sum_{t=1}^{N} \text{SMR}_{jt}$ is the average SMR over all time periods in LAD j. Note that in this study, $N = 67$ denotes the number of weeks. To evaluate temporal autocorrelation for different lags simultaneously, a Ljung-Box test is applied. Following the definition in Section 2.5.2.1, for each area j, the statistic $\tilde{Q}(\hat{r}_j)$ is computed as a weighted sum of the temporal sample correlations $\hat{r}_{jl}$ (from Equation 3.5) up to a lag m, as

$$\tilde{Q}(\hat{r}_j) = n(n+2)\sum_{l=1}^{m}(n-l)^{-1}\hat{r}_{jl}^2, \tag{3.6}$$

which approximately follows a χ^2_m distribution under the assumption of temporal independence. The approximate distribution of the test statistic under the null hypothesis allows testing for temporal autocorrelation up to lag m for each LAD in the study. Since the test for temporal autocorrelation is conducted in 312 LADs simultaneously, a Bonferroni correction is applied, providing a significance threshold of $0.05/312 \approx 0.00016$. Thus, the null hypothesis of no temporal autocorrelation for each LAD is rejected if the Ljung-Box test has a p-value less than 0.00016. When tested for a lag up to $m = 10$, this is true for 310 of the 312 LADs (99.36%), leading to the conclusion that temporal autocorrelation is present in the data.

Having identified spatial and temporal autocorrelation in the COVID-19 mortality data, a spatio-temporal model should be fitted to obtain more accurate estimates of mortality risks that account for the trends and correlations. The specific model used in this study will be introduced next.

3.3 Methodology

Let $j = 1, ..., K(= 312)$ denote the areas (LADs in England) in the study, and let $t = 1, ..., N(= 67)$ denote the time periods (weeks between 1st February 2020 and 14th May 2021). The observed number of deaths in area j and week t is denoted Y_{jt}, and the expected number of deaths is denoted $e_{jt} = e_j$ and does not vary by week.

3.3.1 Data Likelihood

Since the observed number of deaths are counts of rare events, the natural choice is to fit a Poisson log-linear model to the data. Specifically, the model is fitted in a Bayesian setting, where the data likelihood is of the form

$$Y_{jt} \sim \text{Poisson}(e_j \theta_{jt}), \tag{3.7}$$

$$\log(\theta_{jt}) = \beta_0 + \phi_{jt}. \tag{3.8}$$

Here, the goal is to estimate the relative mortality risk θ_{jt} for area j at time t, which is on the same scale as the SMR and the estimated risk can be interpreted as a spatio-temporally smoothed version of the noisy SMR. The natural log of the relative risk θ_{jt} is modelled by an intercept term β_0, and a spatio-temporal trend modelled by random effects $\{\phi_{jt}\}$. In the hierarchical model the mean β_0 is assigned a prior distribution, which will be discussed further in Section 3.3.3. First, the structure of the spatio-temporal random effects is explored.

3.3.2 Spatio-temporal random effects

Section 2.7 introduced several spatio-temporal models for areal data. A common approach of these models is to use a CAR prior to capture the data's spatial structure. Since many of the other models have been pointed out to be potentially over-parameterised, this study implements the spatio-temporal random effects model that was proposed by Rushworth et al. (2014) (recall Section 2.7.4). The joint distribution

of the random effects $(\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N)$ is decomposed as either

$$AR(1): \quad f(\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N) = f(\boldsymbol{\phi}_1) \prod_{t=2}^{N} f(\boldsymbol{\phi}_t | \boldsymbol{\phi}_{t-1}), \tag{3.9}$$

$$\text{or} \quad AR(2): \quad f(\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N) = f(\boldsymbol{\phi}_1) f(\boldsymbol{\phi}_2) \prod_{t=3}^{N} f(\boldsymbol{\phi}_t | \boldsymbol{\phi}_{t-1}, \boldsymbol{\phi}_{t-2}), \tag{3.10}$$

where $\boldsymbol{\phi}_t = (\phi_{1t}, ..., \phi_{Kt})$ denotes the vector of random effects for week t. Note, Rushworth et al. (2014) only consider the AR(1) model but here both the AR(1) and AR(2) models are presented. A spatial prior distribution is assigned to the random effects from the first week (AR(1)) or first and second week (AR(2)), while the conditional distributions of the random effects from the later weeks are assigned spatio-temporal prior distributions as they are assumed to depend on the random effects from previous weeks.

A common method to capture spatial proximity in area-level spatial random effects is through a neighbourhood matrix $\boldsymbol{W}$. In this study, $\boldsymbol{W}$ is a binary adjacency matrix, with the entry in the j-th row and i-th column defined as

$$w_{ji} = \begin{cases} 1, & \text{if LADs } j \text{ and } i \text{ share a border,} \\ 0, & \text{otherwise.} \end{cases} \tag{3.11}$$

The matrix has a null-diagonal, i.e. $w_{jj} = 0$ for $j = 1, ..., K$. Note that other types of neighbourhood matrices could be considered. For example, a particular number of nearest LADs could be considered in the neighbourhood structure, so that LADs could be defined as neighbours even when they do not share a border. Alternatively, distances between centre points of areas could be reflected by weights, representing similarities based on the measured proximities. However, the resulting estimates often end up being reasonably similar for the different neighbourhood matrices, so this study continues with the binary adjacency matrix defined above rather than introducing a more sophisticated neighbourhood structure.

The joint prior distribution of the random effects during week 1 is defined as $\boldsymbol{\phi}_1 = (\phi_{11}, ..., \phi_{K1}) \sim N(\mathbf{0}, \tau^2 \boldsymbol{Q}(\rho, \boldsymbol{W})^{-1})$, where the precision matrix $\boldsymbol{Q}(\rho, \boldsymbol{W})$ is defined as $\boldsymbol{Q}(\rho, \boldsymbol{W}) = \rho(\text{diag}(\boldsymbol{W}\mathbf{1}) - \boldsymbol{W}) + (1 - \rho)\boldsymbol{I}$ with $\boldsymbol{I}$ denoting a $K \times K$ identity

matrix and $\mathbf{1}$ denoting a $K \times 1$ vector of ones. The full conditional distribution of the spatial random effect ϕ_{j1} for area j in week 1, given the spatial random effects of all other areas in that week can then be expressed in closed form:

$$\phi_{j1}|\boldsymbol{\phi}_{-j1} \sim N\left(\frac{\rho\sum_{i=1}^{K} w_{ji}\phi_{i1}}{\rho\sum_{i=1}^{K} w_{ji} + 1 - \rho}, \frac{\tau^2}{\rho\sum_{i=1}^{K} w_{ji} + 1 - \rho}\right), \tag{3.12}$$

where $\boldsymbol{\phi}_{-j1}$ is a vector of spatial effects from week 1 for all areas except for area j. Here, ρ controls the spatial autocorrelation, and τ^2 is a variance parameter. Note that in the case of the AR(2) model, the full conditional distribution of $\phi_{j2}|\boldsymbol{\phi}_{-j2}$ has the same structure, only that the mean depends on the neighbouring spatial effects in week 2, rather than week 1. The prior distribution whose full conditional form is specified in Equation (3.12) is a Leroux CAR prior (recall Section 2.6.2). For $\rho = 1$, it reduces to that of an intrinsic CAR prior (Section 2.6.1), where the mean of the spatial effect in area j during week 1 is completely determined by the spatial effects ϕ_{i1} from the same week, for all areas i that share a border with area j. Further, the variance term in the full conditional distribution decreases as the number of neighbouring areas increases. While $\rho = 1$ suggests strong spatial autocorrelation, $\rho = 0$ implies spatial independence, as the conditional distribution reduces to $\phi_{j1}|\boldsymbol{\phi}_{-j1} \sim N(0, \tau^2)$ for all areas j. The more general form of the full conditional distribution presented in Equation (3.12) allows for greater flexibility than either of the special cases $\rho = 0$ or $\rho = 1$, as the data are used to estimate the strength of spatial autocorrelation controlled by the correlation parameter ρ.

While the full conditional distributions for the spatial prior distributions of the random effects from the first week (AR(1)) or the first and second week (AR(2)) look the same for either version of the model, the full conditional distributions for the spatio-temporal random effects of the later weeks are slightly different. They are defined as:

$$\text{AR(1)}: \quad \boldsymbol{\phi}_t|\boldsymbol{\phi}_{t-1} \sim N\big(\alpha\boldsymbol{\phi}_{t-1}, \tau^2\boldsymbol{Q}(\boldsymbol{W}, \rho)^{-1}\big), \;\; t = 2, ..., N, \tag{3.13}$$

$$\text{or AR(2)}: \quad \boldsymbol{\phi}_t|\boldsymbol{\phi}_{t-1}, \boldsymbol{\phi}_{t-2} \sim N\big(\alpha_1\boldsymbol{\phi}_{t-1} + \alpha_2\boldsymbol{\phi}_{t-2}, \tau^2\boldsymbol{Q}(\boldsymbol{W}, \rho)^{-1}\big), \;\; t = 3, ..., N, \tag{3.14}$$

where α in the AR(1) model controls the level of temporal autocorrelation, such that $\alpha = 0$ indicates temporal independence, while $\alpha = 1$ indicates strong temporal autocorrelation and makes the distribution a first order random walk. For the AR(2) model α_1 and α_2 are again temporal dependence parameters that determine the relationship between spatio-temporal random effects that are temporal neighbours of order 1 and 2, respectively. Note that as a special case, a second order random walk can be obtained for $\alpha_1 = 2$ and $\alpha_2 = -1$.

3.3.3 Hierarchical specifications

The prior distribution of the mean of the log-transformed mortality risk θ_{kt} from Equation (3.8) is

$$\beta_0 \sim N(\mu_0 = 0, \sigma_0^2 = 10,000), \tag{3.15}$$

which is a weakly informative prior distribution, reflecting that no prior knowledge is available for the underlying mortality risk of COVID-19 in England in the given time frame. The hyperpriors for the parameters (ρ, τ^2, α) or $(\rho, \tau^2, \alpha_1, \alpha_2)$ from the spatio-temporal random effects models are given by:

$$\begin{aligned} \text{AR(1)}: \quad \rho, \alpha &\sim \text{Uniform}(0,1), \\ \tau^2 &\sim \text{Inverse-Gamma}(a = 1, b = 0.01). \end{aligned} \tag{3.16}$$

$$\begin{aligned} \text{AR(2)}: \quad \rho &\sim \text{Uniform}(0,1), \\ f(\alpha_1, \alpha_2) &\propto 1, \\ \tau^2 &\sim \text{Inverse-Gamma}(a = 1, b = 0.01). \end{aligned} \tag{3.17}$$

Thus for both AR(1) or AR(2) models, the spatial and temporal dependence parameters ρ and α or (α_1, α_2) are assigned flat (non-informative) priors, and the variance parameter τ^2 from the spatial random effects is assigned a conjugate inverse-gamma prior with hyperparameters $a = 1$ and $b = 0.01$. These are the default values in the function `ST.CARar` from the package `CARBayesST` (Lee et al., 2021), which is used to fit the Bayesian hierarchical model in `R`, via MCMC simulation.

3.4 Results

The model proposed by Rushworth et al. (2014) outlined in Section 3.3 is fitted to the COVID-19 mortality data for both the AR(1) and AR(2) versions. Estimated risks are obtained from the posterior mean of the fitted values divided by the expected counts. For each version, the MCMC algorithm produces 2,200,000 simulations for each parameter in the model, the first 200,000 of which are discarded as burn-in period. The simulations are thinned by saving only every 1,000-th simulation to reduce the autocorrelation in the Markov chains. Ultimately, this provides 2,000 simulated values for each parameter.

Geweke diagnostics (recall Section 2.3.2.3) between (-2,2) suggest no evidence of lack of convergence in the algorithm for either version. For example, for the AR(2) model the Geweke diagnostics take on the values 1.4 for β_0, -0.8 for τ^2, -0.5 for ρ, 0.0 for α_1, and -0.2 for α_2. Trace plots are examined for both models, and they do not show strong evidence against convergence. The trace plots for the AR(2) model are presented in Section A.1 of the appendix. The Deviance Information Criterion (DIC, recall Section 2.4) is used to compare the two models. The AR(1) model's DIC value is 69,067, and the AR(2) model's DIC value is 68,772. Hence, the model with second-order temporal autocorrelation fits the data slightly better and is used to estimate the relative risk for each LAD and week. Posterior predictive checks are carried out to confirm that the AR(2) model fits the data appropriately, and these can be found in Section A.2 of the appendix. Additionally, an analysis of the sensitivity of the results to the prior choice for the variance parameter is provided in Section A.3.

The remainder of this section answers the central questions of interest as follows.

- **Section** 3.4.1:

    1. How long after the implementation of lockdown did mortality risks reduce at a national level, and did this vary by lockdown?

- **Section** 3.4.2:

    2. How did the temporal trends in mortality risks differ by region in England in the weeks following the implementation of lockdown?

3. Which local authorities were exposed to the highest average risks in those weeks?

- **Section** 3.4.3:

4. Which local authorities shared similar temporal trends in mortality risks?

3.4.1 How long after the implementation of lockdown did mortality risks reduce at a national level, and did this vary by lockdown?

The first question is the most important from an epidemiological perspective in this study because it quantifies how long lockdowns have to be in place before mortality risks reduce. It can be answered by comparing the panels in Figure 3.5, which show boxplots of the distributions of the posterior mean estimated mortality risks across all LADs in the weeks preceding, during, and after each of the three lockdowns (left), and plots with 95% credible intervals around the median average estimated risk across England (right). Panels (a), (b), and (c) show these results for the first, second, and third lockdown, respectively, and all the risks presented are relative to the average risk across England for the entire study period. In each panel, week -1 is the week preceding the lockdown, week 0 is the onset week of lockdown, week 1 is the first week after the onset of lockdown, and so on. The weeks coloured in beige comprise the lockdowns. The y-axes measuring estimated risk and average estimated risk are on the same scale for the three panels to allow comparison across the lockdowns. The blue dashed line shows the median estimated risk across England from the last week before lockdown, and the red dashed line represents a risk of 1. Note that week 7 in panel (b) is the same as week -1 in panel (c).

The weeks after each lockdown are included to see if the lockdowns had a lasting impact on mortality risk. In an attempt to analyse risk over similar periods, the panels for the first and third lockdowns were each set to contain thirteen weeks from the start of the lockdown. However, the panel for the second lockdown ranges over fewer weeks, as week 7 is already the last week before the third lockdown was introduced. The y-axis measuring estimated risk is on the same scale for the three

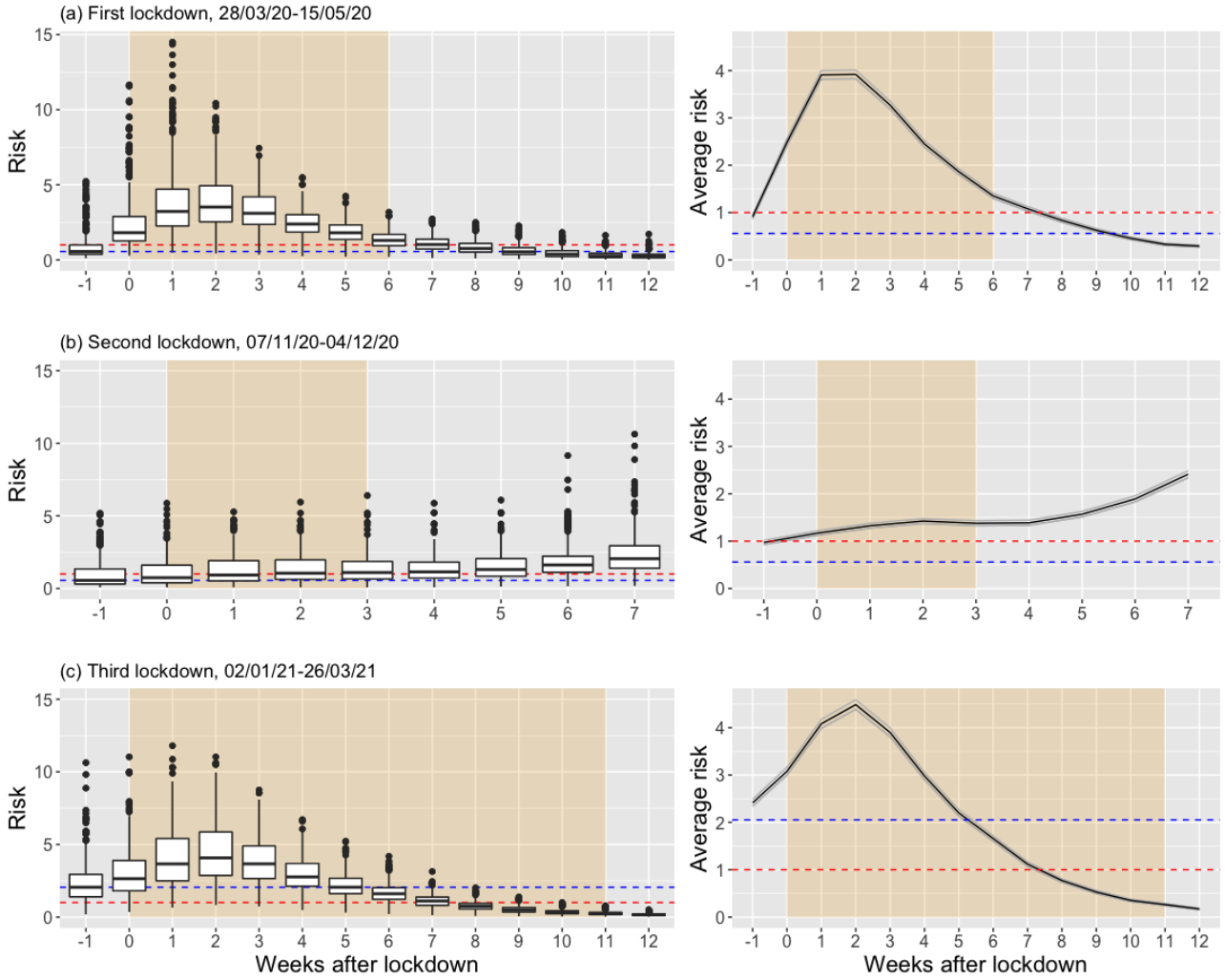


Figure 3.5: Left: Boxplots of the posterior mean estimated risks across LADs, by week after lockdown. Right: Median line with 95% credible intervals for the average estimated risk across LADs, by week after lockdown. The weeks of national lockdown are highlighted in beige. The red dashed lines represent a risk of 1. The blue dashed lines show the median estimated risk across England from the last week before lockdown.

panels, allowing comparison across the lockdowns. The red dashed lines represent a risk of 1 in all three panels. Hence, any risks above these lines are elevated compared to the average risk over all LADs and weeks of the study. Finally, the blue dashed lines show the median estimated risks across England for the week preceding each lockdown, making it easier to identify how long it took for the estimated risk to reduce to pre-lockdown levels. The plots on the right were included to visualise the variability in the accumulated risks by week across the entire country. Since the 95% credible intervals are very narrow, there does not appear to be much variation in the averages of the simulated estimated risks across all LADs.

The first striking feature apparent from Figure 3.5 is that the second lockdown (in panel (b)) exhibits a very different trend in risk over time compared with the first and third lockdowns. The second lockdown shows a constant or slightly increasing trend

in risk throughout, while the overall risk in the first and third lockdowns shows an increasing trend in the first three weeks of lockdown, followed by a decreasing trend. The second lockdown was much shorter, lasting only four weeks, compared to seven weeks of lockdown 1 and twelve weeks of lockdown 3. The short duration of lockdown 2 may be part of the reason for its different temporal trends in risk, as it was not in place long enough to have a sizeable effect in reducing risk. Additionally, the lack of an increasing trend in the first few weeks after the introduction of lockdown 2 may suggest that the pandemic was not yet approaching a severe new wave in terms of mortality. A possible reason for the lack of an increasing trend could be that England had implemented a three-tier system of mobility restrictions that started on 14th October before lockdown 2, with varying degrees of restrictions for the LADs in the three tiers, which Davies et al. (2021) suggested has had a sizeable effect in reducing the number of COVID-19 deaths.

The second key feature from Figure 3.5 is that for lockdowns 1 and 3, the mortality risk increased in the first three weeks after the introduction of the lockdown before it started to reduce from the fourth week onward. The reason for this is the lag between a COVID-19 infection and mortality, which the ONS estimate is between 21 and 25 days on average (ONS, 2021a), which corresponds to between 3 and 3.5 weeks. Thus the high number of infections in the last few weeks before the lockdown would transfer to the high mortality risks observed three weeks into lockdown. The sharp increase in mortality risks in the first weeks of lockdowns 1 and 3 suggests that they might have been implemented too late, as it could have been beneficial to stop the increase earlier.

The final important finding from Figure 3.5 is the time it took for the risks to reduce to baseline levels. Taking this baseline level as a risk of 1, it took eight weeks after the introduction of both lockdowns 1 and 3 for the median risk across England to reduce to that level. In the previous week 6, the risks were still largely above this level, with 73.4% (lockdown 1) and 85.3% (lockdown 3) of LADs having risks above 1. However, this is a slightly unfair comparison because the median risks across England were 0.56 and 2.05 in the week preceding lockdowns 1 and 3, respectively. The blue dashed lines in the figure show the median risks in the week preceding each

lockdown, which can be considered an alternative baseline level. Therefore, the figure reveals that it took ten (lockdown 1) and six (lockdown 3) weeks after introducing the lockdowns for the median risks across England to reduce back to these pre-lockdown levels. This large difference is because the median risk before the introduction of lockdown 1 was much lower than that of lockdown 3. Note that the median risk across England rose to around 3.5 (lockdown 1) and 4.1 (lockdown 3) in week 3 after each lockdown. Hence, reducing to a risk of 2.05 (the risk before lockdown 3) from 4.1 was much easier than reducing to 0.56 (the risk before lockdown 1) from 3.5. Note that lockdown 1 was lifted in its seventh week, while lockdown 3 was lifted in its twelfth week. Nonetheless, their temporal trends in risks were strikingly similar. However, England was nearing the summer when lockdown 1 was lifted. In contrast, lockdown 3 only lasted until the end of March, which should be considered as the number of deaths tends to be lower in the warmer months. There might be other reasons for the longer duration of lockdown 3, and Section 3.5 provides a further discussion on this topic.

As a conclusion for the first main question, lockdowns may reduce mortality risk after approximately four to five weeks. However, restrictions may need to be kept in place for additional time to reduce risk far enough to avoid an immediate increase after the restrictions are lifted, as was observed for the second lockdown.

3.4.2 How did the temporal trends in mortality risks differ by region in England, and which LADs were at the highest risk?

The previous section explored the national mortality risk distribution trends in the weeks preceding, during, and after the three national lockdowns in England. This section looks at spatial patterns by analysing geographical differences in the temporal risk trends at a regional level, and identifies the local authority districts (LADs, the areal units in the data) with the highest average mortality risks in the weeks after lockdown. Recall that there are 312 LADs which are exactly nested within the nine English regions (East of England, East Midlands, London, North East, North West,

South East, South West, West Midlands, and Yorkshire and The Humber), as shown in Figure 3.1.

The analysis at the larger region scale uses weekly population-based weighted average estimated risks and scaled estimated risks. The weekly weighted average risks $\hat{\theta}_{rt}$ for region r and week t are computed as

$$\hat{\theta}_{rt} = \frac{1}{P_r} \sum_{j \in r} P_j \times \hat{\theta}_{jt}, \tag{3.18}$$

where $\hat{\theta}_{jt}$ denotes the risk from area j and week t, and $j \in r$ indicates that LAD j falls into region r, while P_j and P_r denote the population sizes of LAD j and region r, respectively. Weighted averages are taken with regard to population size as the risk in a more populous LAD will have a more substantial impact on the region's risk than the risk in a less populous LAD. Since the LADs are not large, the accumulations at a regional level can be used to see if there are any larger geographical temporal risk trends.

The weekly estimated risk in each LAD is divided by its estimated risk from the week before the respective lockdown was implemented to obtain scaled estimated risks. That is, the scaled estimated risk $\hat{s}_{jl}$ is computed as

$$\hat{s}_{jl} = \frac{\hat{\theta}_{jl}}{\hat{\theta}_{j(-1)}}, \qquad l = -1, 0, 1, 2, \ldots, \tag{3.19}$$

where $\hat{\theta}_{jl}$ denotes the estimated risk from LAD j and week l after the introduction of lockdown. So, as before, $l = -1$ denotes the week before the introduction of lockdown, $l = 0$ denotes the week during which the lockdown was introduced, $l = 1$ denotes the first week after the introduction of lockdown, and so forth. From the scaled estimated risk, weighted averages $\hat{s}_{rl}$ for region r and week l after the introduction of lockdown are computed as

$$\hat{s}_{rl} = \frac{1}{P_r} \sum_{j \in r} P_j \times \hat{s}_{jl}, \tag{3.20}$$

with notation analogous to that of Equation (3.18). Thus, the scaled estimated risk should be interpreted as the risk from a specific week during or after lockdown

relative to the risk in the week preceding lockdown. For example, if the scaled risk of a particular week is 2.5, then this suggests that risk was 2.5 times as large in that week, relative to the week preceding lockdown.

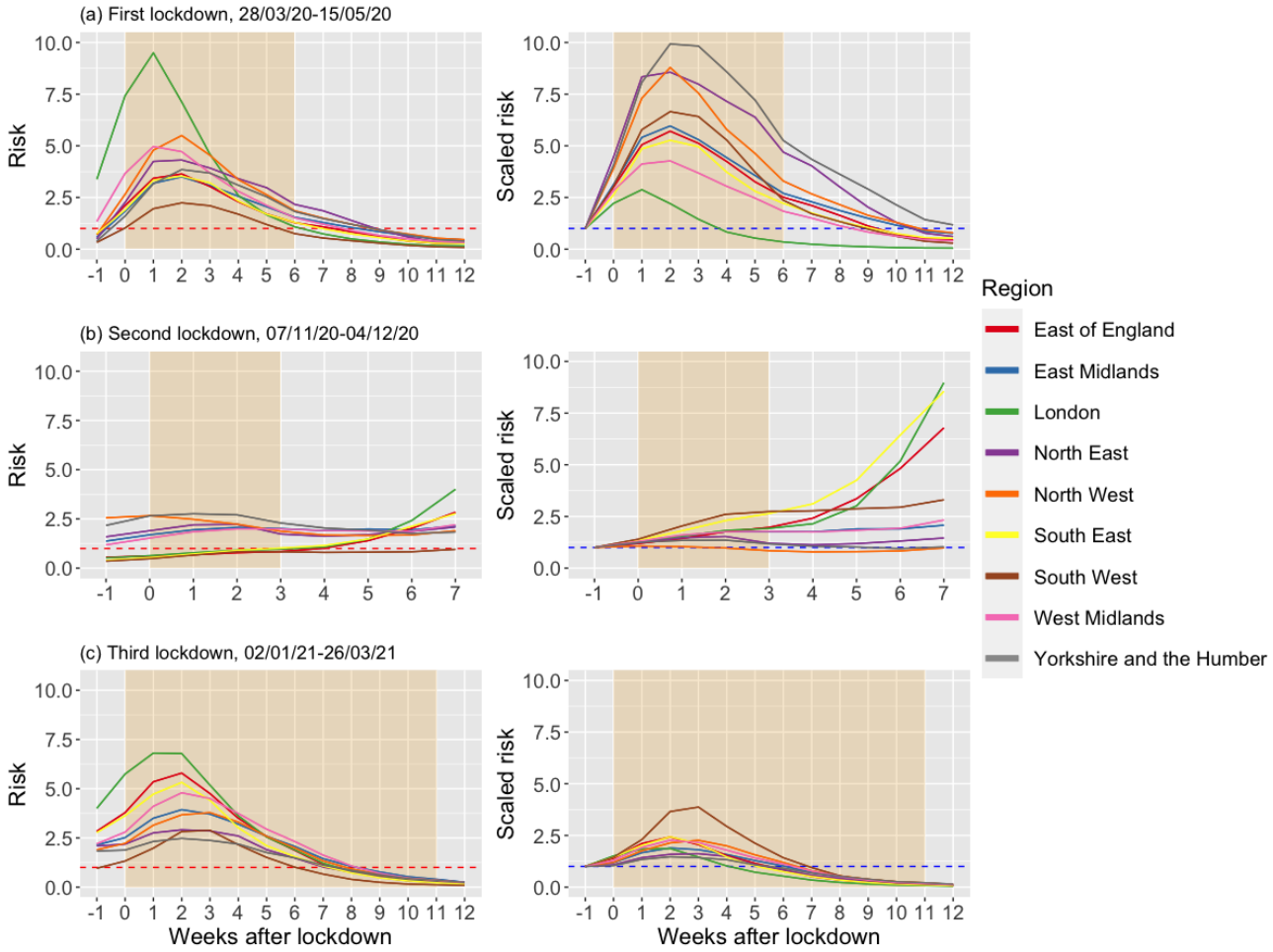


Figure 3.6: Average estimated risk (left) and average scaled estimated risk (right) in weeks following the implementation of national lockdown, by region. The weeks of national lockdown are highlighted in beige. The red dashed lines represent a risk of 1. The blue dashed lines represent a scaled risk of 1.

Figure 3.6 shows line plots for the population-based weighted averages of estimated risk (left) and scaled estimated risk (right) by week and region. As before, the red dashed lines in the plots on the left represent a risk of 1. The blue dashed lines in the plots on the right represent a scaled risk of 1. The number of weeks that elapse before the line from each region moves across the red and blue lines provides measures for lockdown effectiveness to stop and reverse the trend of increasing mortality risk at a regional level. Note that week 7 in panel (b) is again the same as week 0 in panel (c).

First, note that the plots confirm the observation from the previous section that the temporal trends in risks in lockdown 2 were very different from those in lockdowns

1 and 3. Recall that lockdown 2 featured relatively constant risk over time at a national level. At a regional level, some differences are apparent. While the mortality risks stayed relatively constant for all regions when the lockdown was in place, the risks increased rapidly in London, South East, and East of England after the lockdown was lifted. This feature is particularly prominent on the scaled estimated risk plot. Since the initial risks were very low in these regions, the increased estimated risks in the later weeks are amplified when displayed as scaled risks, clearly separating the temporal trends in risks in the southeast from the remaining regions of England. Note that the Alpha variant of COVID-19 was detected in the Kent area (South East) in September 2020, and this variant was estimated to be 1.5 times more transmissible than earlier variants of COVID-19 (Page and McNamara, 2021). Furthermore, the mortality risk for those infected with the Alpha variant was estimated to be 1.6 times higher than those with earlier variants. Hence, it is likely that the Alpha variant of COVID-19 drove the distinct mortality risk trends in southeast England after lockdown 2. Since the main interest of this study is on reductions in mortality risk following the implementation of lockdown, the remainder of this study will focus on lockdowns 1 and 3, as for lockdown 2 no reduction in mortality risk was observed at a national or regional level.

First, consider the trends in the regional average estimated risks (shown on the plots on the left in Figure 3.6). Recall that for lockdowns 1 and 3, the temporal risk trends were fairly similar at a national level, as the risks increased in the first three weeks of both lockdowns before they started to decrease. This general trend can be observed at a regional level as well. The most striking feature is that the average estimated risk was highest in London during the first three weeks of both lockdowns, with an especially high peak in the second week of lockdown 1. The risk started to decrease in all regions sometime during the third or fourth week of lockdown 1, while it decreased only in the fourth or fifth week of lockdown 3. The risk in London reduced quickly after implementing each lockdown, which is especially noticeable in the plot for lockdown 1. London was the second region to reach an estimated risk of 1 (red line) in the seventh week of lockdown 1. According to Batty et al. (2021), the proportion of essential workers in the UK is approximately 23.6%, while that of

London is only 16%. Hence, the smaller proportion of essential workers could explain why mortality risks reduced so quickly in London after the first weeks of lockdowns 1 and 3, while they reduced more slowly for some of the other regions. A possible reason for this hypothesis is that while a lockdown is in place, essential workers have much higher mobility and face-to-face interactions than non-essential workers who are ordered to stay at home and only go out for physical exercise or essential shopping. Generally, it seems plausible that reduced mobility has a greater impact in urban areas, compared to rural areas where the number of face-to-face interactions might be lower by default.

For lockdowns 1 and 3, the South West was the first region to reach an estimated risk of 1 in the fifth and sixth week of lockdown, respectively. All regions reached an estimated risk of 1 in the seventh or eighth week after the introduction of lockdown 3. In contrast, for lockdown 1 some regions only reached that level in the tenth week after the introduction of lockdown. Thus, although risks started to reduce later in lockdown 3, they appear to have reduced more quickly than in lockdown 1. Part of the reason could be that lockdown 3 was lifted later than lockdown 1, which was lifted before most of the LADs reached a risk of 1.

The plots on the right, showing the scaled estimated risks, provide additional insights into risk trends. Note that the risk in London did not increase as drastically during lockdown 1, considering the high level where it started. Further, the risk in London was quickest to return to its initial level within five weeks of lockdown 1. Relative to their lower initial risk levels, the risk increased most drastically in Yorkshire and The Humber, North East, and North West. In these regions, the estimated risks also took the longest time to return to their initial levels. While this might suggest a less effective lockdown in these regions, note that this trend did not repeat in lockdown 3. Here, the only region that stands out is the South West, which started at a much lower risk level than the other regions, followed by a rapid increase in risk during the first three weeks of the lockdown. However, the plot should not stir up too much concern; it may have taken the South West the longest to return to its initial (particularly low) level of risk, yet it was the first region for which the estimated risk fell below 1 (left plot). The scaled risk plot emphasises that the temporal trends in

risks were very similar at a regional level during the third lockdown for the remaining eight regions.

In conclusion to the second main question, regional differences in the temporal risk trends have varied for the three lockdowns. In the first lockdown, it took longer for risks to reduce in the northern regions of England (North East, North West, Yorkshire and The Humber) compared to the rest of the country. In the second lockdown, the southern regions (London, South East, South West, and East of England) started at much lower risk levels than the other regions. London, South East, and East of England had the steepest risk increase after the lockdown was lifted. In the weeks during and after the second lockdown, there was no substantial reduction in mortality risks for any region, so the second lockdown will not be analysed further in the remainder of this study. There had been no substantial differences in temporal risk trends at a regional level in the third lockdown. Overall, although some regional differences in risk trends are apparent during the lockdowns, there was no clear hierarchy for the regions of England in temporal risk trends across the three lockdowns.

The third main question asks for the LADs with the highest averages in mortality risk in the weeks following the introduction of the lockdown. Table 3.1 presents the five LADs with the highest average risks for the first and third lockdowns. For both lockdowns, the average risks and standard deviations are computed over the fourteen weeks displayed in Figure 3.6. The high average risks suggest that these LADs were particularly vulnerable, and the standard deviations are provided to check if any of these LADs had extreme changes in mortality risk, which could be caused by a particularly high peak in risk followed by a rapid decrease.

Before discussing the information provided in Table 3.1, note that COVID-19 mortality risks had previously been linked to deprivation, such that more deprived areas tend to be exposed to greater risks (Tinson, 2021). The government website (Ministry of Housing, 2019a) provides a 2019 ranking of the index of multiple deprivations (IMD) for lower layer super output areas (LSOAs), which combines deprivation scores from several domains (Ministry of Housing, 2019c provides an infographic for the IMD). They also provide population-based weighted averages of LSOA ranks for each LAD (Ministry of Housing, 2019b), from which a deprivation ranking for the

LADs in the COVID-19 mortality data is created. Now, Table 3.1 can be viewed under consideration of this deprivation ranking and the regional trends discussed earlier in this section.

Lockdown	LAD (IMD rank)	Average risk	SD	Region
1	Lambeth (42)	3.46	3.77	London
	Hackney (7)	3.49	4.33	London
	Haringey (37)	3.49	4.27	London
	Newham (12)	3.83	4.70	London
	Brent (49)	4.10	4.80	London
3	Castle Point (181)	4.05	3.70	East of England
	Tower Hamlets (27)	4.11	3.94	London
	Slough (73)	4.12	3.21	South East
	Barking and Dagenham (5)	4.29	4.07	London
	Newham (12)	4.65	4.29	London

Table 3.1: Five LADs with highest average risk across time for lockdowns 1 and 3.

Table 3.1 shows the five LADs with highest average risk across time for lockdowns 1 and 3, and the national IMD rank for the 312 LADs in this study is provided for each LAD (in parentheses). Here, a low rank implies high deprivation (i.e., the LAD with IMD rank 1 is the most deprived LAD). The table shows that the five LADs with the highest average risks during lockdown 1 are all in London. Recall that according to Figure 3.6, London was the region with the highest peak risk, which could explain a high average risk for at least some of its LADs. The peak in the region as a whole was followed by a rapid decrease in risk, which also seems to be the case for the five LADs since the standard deviations are larger for the LADs with higher values of average risk. The five LADs with the highest average risk during lockdown 3 are in London, South East, and East of England. Again, these regions had the highest peaks in estimated risk during that lockdown, as shown in Figure 3.6. Amongst the five LADs in the table for lockdown 3, the standard deviations are again larger for the LADs with higher average values, except for Slough, which has the third highest average risk but the lowest standard deviation of the five LADs. Taking a closer look at the data confirms that the lower standard deviation is due to the risks having decreased much slower in Slough than in the other LADs in the table.

According to the previously described deprivation ranking, out of the 312 LADs in mainland England, the five LADs with the highest average risk in lockdown 1 (Lambeth, Hackney, Haringey, Newham, and Brent) are amongst the 50 LADs with

the highest deprivation. Further, they are amongst the 11 most deprived districts of the 33 LADs in London. For lockdown 3, the LADs with the highest average risk that are in London are Barking and Dagenham, Newham, and Tower Hamlets, which place first, third, and fourth in the ranking of highest deprivation amongst LADs in London. Slough has the sixth highest deprivation amongst the 63 LADs in South East, and Castle Point has the 21-st highest deprivation amongst the 45 LADs in East of England. While the deprivation in Castle Point does not seem exceptionally high compared to the other local authorities in East of England, the other four LADs have some of the highest deprivation in their respective regions.

Thus, the observations from answering the third main question lead to a hypothesis that at least two driving factors cause differences in risk trends at a local authority level. The first is the underlying regional trend, as the five LADs with the highest average risk in each of the lockdowns all belong to the regions that have experienced the highest levels of risk. Secondly, most LADs with the highest average risk belong to the most deprived local authorities in their respective regions. Hence, the LADs at the highest risk appear to be deprived districts within regions at higher risk.

3.4.3 Which local authorities shared similar temporal trends in mortality risks?

The fourth and final main question requires checking if there are groups of LADs with similar temporal risk trends, considering the first and third lockdowns separately. Non-hierarchical clustering methods can be used to identify these groups. Specifically, k-means clustering (recall Section 2.8.1.1) is applied for each lockdown such that the LADs are assigned to between one and ten clusters according to (dis)similarities in temporal trends in risks. Note that it is common to standardise the observations from each variable before applying k-means clustering. However, here, the goal is to distinguish LADs both by peak risk and time period of when the peak risk occurred, so the estimated risks are not standardised for each week so that the k-means algorithm can pick up these differences. Furthermore, it is desirable to assign clusterings based on the same number of weeks for lockdowns 1 and 3 to compare temporal trends in

risks for the two lockdowns. Recall that lockdown 3 was only lifted in the twelfth week after the introduction of lockdown, while lockdown 1 was already lifted in the seventh week. Hence, at least the first twelve weeks after the introduction of lockdown should be considered to include all weeks of lockdown 3, while not too many more weeks should be included since lockdown 1 had already been lifted in the seventh week. With that in mind, clusterings are assigned over fourteen weeks ranging from the week before each lockdown to the thirteenth week after the introduction of the lockdown. The within-cluster sum of squares and average silhouette width plots (recall Section 2.8.1.3) are used to verify that clustering is appropriate for the data at hand and to identify the optimal number of clusters, respectively. The average temporal trends in risks for the resulting clusters are then presented in Figure 3.8, and the cluster assignments for the LADs are displayed on maps in Figure 3.9, allowing a discussion of geographical differences for temporal trends in risks from the first and third lockdowns.

Note that possibly outlying LADs regarding the temporal trends in risks are identified using the single linkage method from hierarchical clustering (recall Section 2.8.2). The k-means algorithm is performed on reduced datasets without these potential outliers to check the sensitivity of the k-means algorithm for the studied mortality data. The resulting clusterings are very similar for the full and reduced datasets, suggesting that the clustering obtained for the full dataset is not affected by extreme cases. Thus, the results for the full dataset are presented below.

Figure 3.7 shows the plots for average silhouette width and the within-cluster sum of squares (recall Section 2.8.1.3) for clusterings obtained from the k-means algorithm applied to the estimated risks in the before mentioned weeks from the first and third lockdowns, for $k = 1, 2, \ldots, 10$. There is a substantial drop in the within-cluster sum of squares when moving from one cluster to two clusters, which suggests that clustering is meaningful. For both lockdowns, the average silhouette width is greatest for two clusters. Hence, the LADs are assigned to two clusters for the first and third lockdowns.

For the last week before each lockdown and the thirteen weeks after the introduction of lockdown, weekly weighted averages of estimated risk are computed by

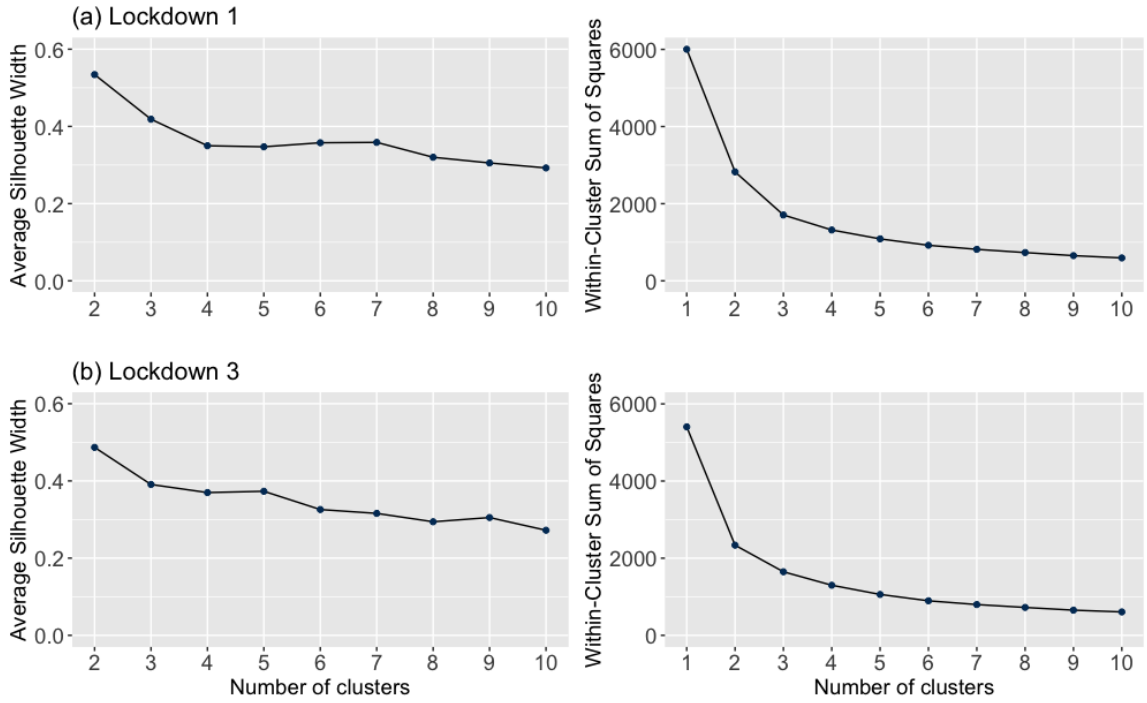


Figure 3.7: Average silhouette width (left) and within-cluster sum of squares (right) for the clusterings of the estimated risk by lockdown, obtained for different numbers of clusters.

cluster. That is, the weighted average estimated risk $\hat{\theta}_{ct}$ for cluster c and week t is computed as

$$\hat{\theta}_{ct} = \frac{1}{P_c} \sum_{j \in c} P_j \times \hat{\theta}_{jt}, \tag{3.21}$$

where P_c and P_j are the population sizes for cluster c and LAD j respectively, and $j \in c$ denotes that LAD j is in cluster c. The risks are accumulated by cluster to obtain summary statistics, since it is not practical to look at the risks of all 312 LADs simultaneously. The averages are weighted by population size since the risks from the LADs with larger populations contribute stronger to the risk of the entire population in that cluster than the risks from LADs with smaller populations.

Figure 3.8 presents plots that show the median average estimated risks with 95% credible intervals by cluster and week, over the 2,000 simulations obtained from the fitted model. The red dashed line indicates an estimated risk of 1. The 95% credible intervals are barely visible, suggesting that there is very little variation in the simulated average risks by cluster. The trends in risk look relatively similar for cluster 1 from each lockdown. The same holds for cluster 2. In the first three to four weeks of each lockdown, the LADs in cluster 1 were at a substantially higher average risk

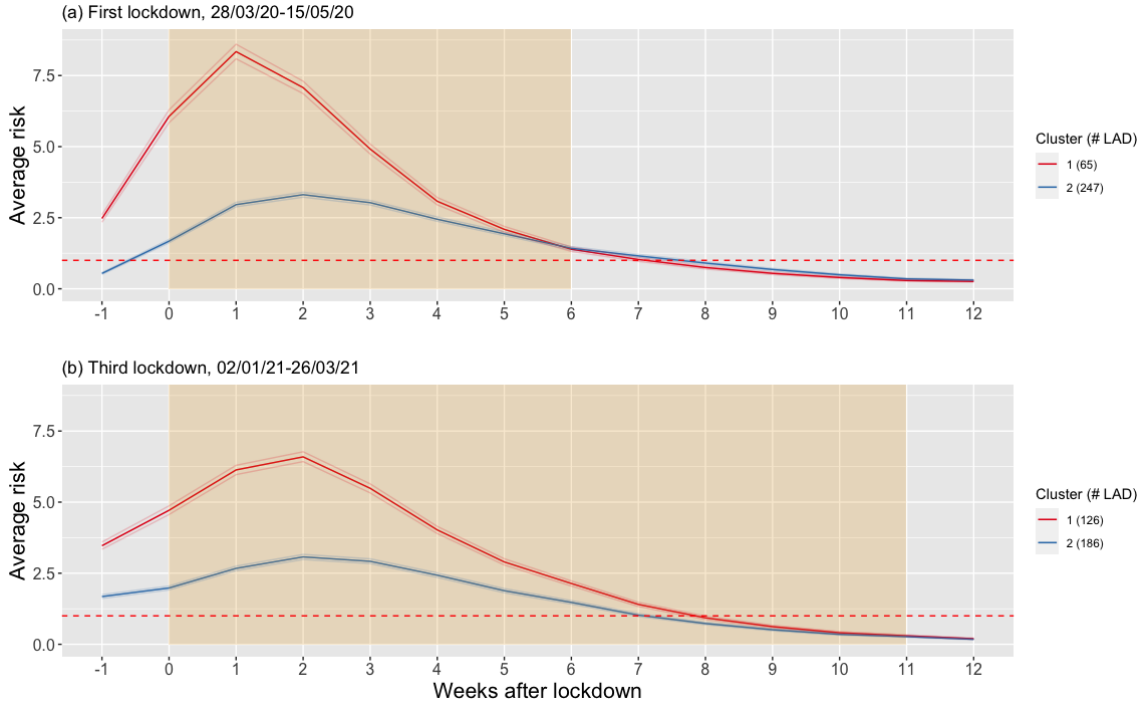


Figure 3.8: Median average estimated risks with 95% credible intervals in the weeks following the implementation of national lockdown, by cluster. The weeks of national lockdown are highlighted in beige.

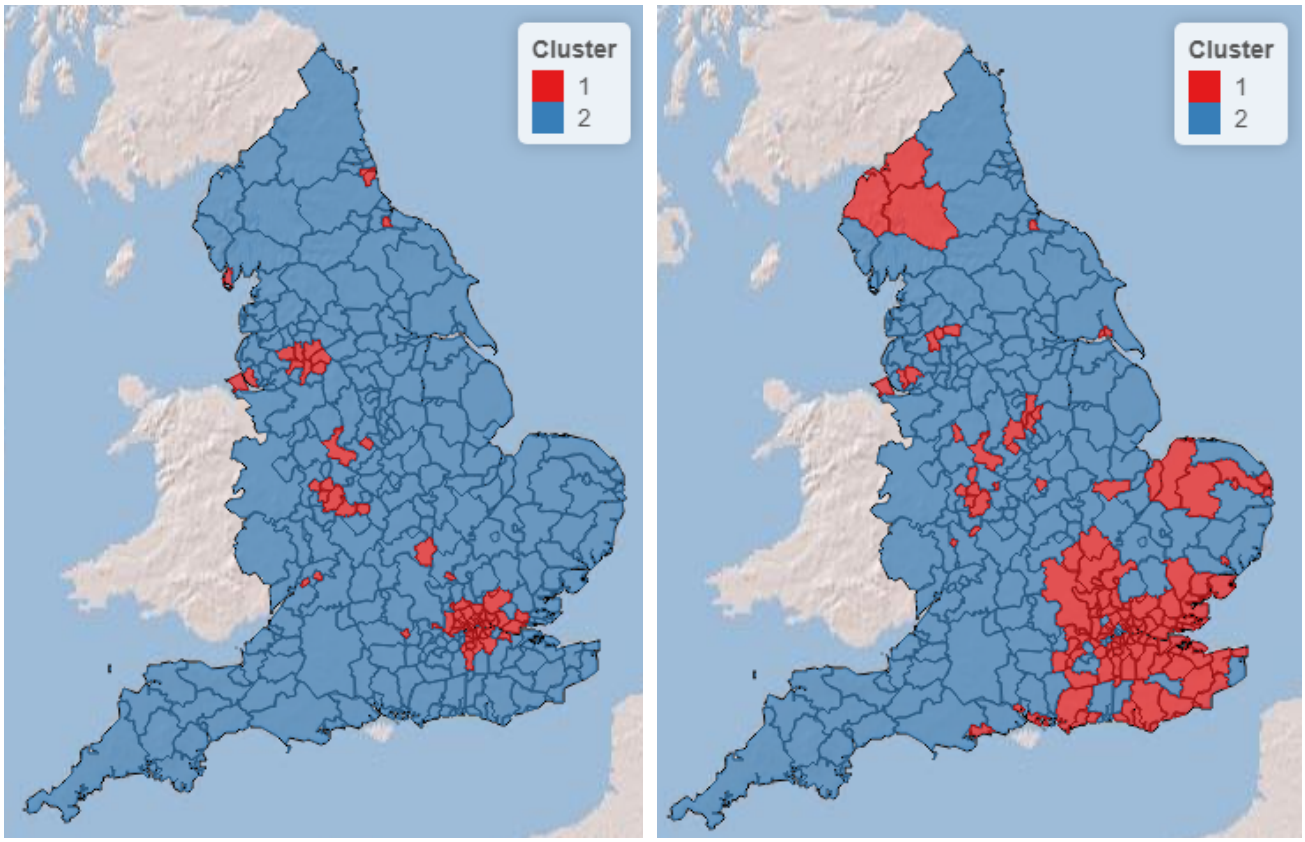
than those in cluster 2. The peak average estimated risk for cluster 1 of lockdown 3 was not as high (approximately 6.59) as the peak average estimated risk for cluster 1 of lockdown 1 (approximately 8.34). However, note that cluster 1 of lockdown 1 consists of only 65 LADs, while cluster 1 of lockdown 3 consists of 126 LADs. The peak average estimated risk for cluster 2 was similar for the two lockdowns (approximately 3.31 for lockdown 1 and 3.07 for lockdown 3). Lastly, the average estimated risk from the two clusters became relatively similar towards the last weeks for both lockdowns, from the sixth week of lockdown 1 and the tenth week of lockdown 3 onward.

	Lockdown 3	
Lockdown 1	Cluster 1	Cluster 2
Cluster 1	45	20
Cluster 2	81	166
	Rand index: 0.5607	

Table 3.2: Contingency table for the cluster memberships of LADs in the first and third lockdowns.

Table 3.2 shows a contingency table that provides additional insights into the cluster assignments for the first and third lockdowns. It indicates that 45 of the 65 LADs in cluster 1 of the first lockdown were also in cluster 1 of the third lockdown, while 81 LADs switched from cluster 2 of lockdown 1 to cluster 1 of lockdown 3. Thus, most of the LADs at higher risk in lockdown 1 were also at higher risk in lockdown 3,

while many additional LADs from cluster 2 in lockdown 1 were assigned to cluster 1 of lockdown 3. The Rand index (recall Section 2.8.1.4) of 0.5607 between lockdowns 1 and 3 suggests that the two clusterings are not very similar. However, the difference between the two clusterings appears to be mostly driven by the much larger number of LADs in cluster 1 of lockdown 3 (126) than in cluster 1 of lockdown 1 (65).



(a) First lockdown **(b)** Third lockdown

Figure 3.9: Maps showing clusters that were formed based on the temporal trends in the weekly estimated risks, by lockdown.

The ultimate goal for this part of the study is to see any geographical patterns for LADs with similar temporal risk trends. Therefore, Figure 3.9 shows maps for lockdowns 1 and 3 that display the cluster memberships of the LADs in the two lockdowns.

In lockdown 1, most of the separation into the two clusters appears to be explained by an urban/rural divide. The peak estimated risk was higher in LADs close to London (9.3 million people), Manchester (2.73 million people), Birmingham (2.6 million people), Liverpool (902,000 people), and Sunderland (341,000 people), while lower in most rural areas (the approximate population sizes in 2020 were obtained from Statista, 2020). In lockdown 3, some urban/rural divide is still noticeable, as

some LADs in the higher risk cluster are close to London, Birmingham, Liverpool, and Leicester (552,000 people) and to smaller urban areas such as Hull or Worcester. However, the most prominent feature is that most LADs at higher risk levels are in the southeast of England in London (82% of LADs in cluster 1), East of England (71% of LADs in cluster 1), and South East (62% of LADs in cluster 1) which might be explained by the more accelerated spreading of the Alpha variant of COVID-19 which originated in Kent in the southeast. Additionally, Grint et al. (2021) suggest that 11% of Alpha variant cases of COVID-19 from their study population of 185,234 people who tested positive between 16th November 2020 and 11th January 2021 were observed in the North West, which could explain the higher risks in the adjacent LADs of Eden, Allerdale, and Carlisle (the three large adjacent areas in cluster 1 in the North West).

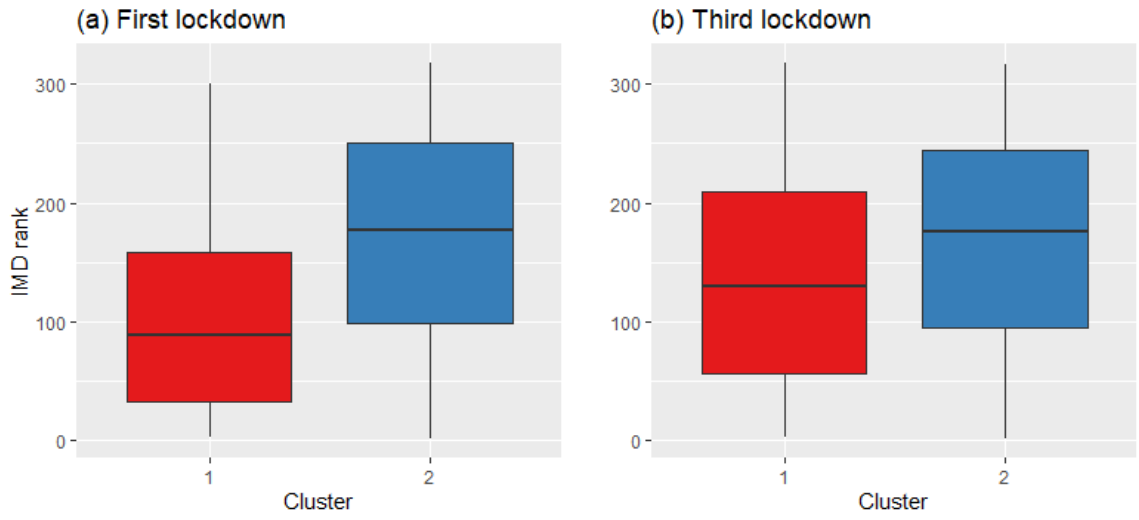


Figure 3.10: Boxplots for the IMD ranks of LADs in the two clusters of lockdowns 1 and 3.

Figure 3.10 shows the distribution of IMD rankings in each cluster for lockdowns 1 and 3 (where a lower rank implies higher deprivation, e.g., the area ranking first has the highest deprivation). Cluster 1 of lockdown 1 has a median IMD rank of 89, while that of cluster 2 is 176. In comparison to the median ranks from lockdown 1, the median IMD rank in cluster 1 of lockdown 3 is substantially higher at 130, while that of cluster 2 is the same at 176. Hence, there might be an association between higher deprivation and increased risks during lockdown 1. Given the urban/rural divide observed for lockdown 1, this suggests that more deprived areas within urbanisation

have been at greatest risk in lockdown 1. In contrast, in lockdown 3, most of the divide into the two clusters seems to be explained by the spreading of the Alpha variant.

3.5 Discussion

The motivation for this study was to explore the temporal trends in COVID-19 mortality risks after national lockdowns, with the main quantity of interest being how long it took for them to effectively reduce mortality risks below the levels from before the lockdowns were introduced. The study was based in England. The temporal trends in mortality risks following the three national lockdowns were investigated using weekly mortality counts on a medium area scale of local authority districts (LADs). A Poisson log-linear model was fitted to these count data in a Bayesian setting using MCMC methods, where the spatio-temporal trends in mortality risks were modelled with a spatially correlated second-order autoregressive process. The analysis highlighted some key findings, which are summarised below.

At a national level (see Section 3.4.1), for lockdowns 1 and 3, it took around three weeks after the introduction of lockdown for mortality risks to stop increasing and ten weeks (lockdown 1) and six weeks (lockdown 3) respectively for the risks to reduce to pre-lockdown levels, which could have been due the differences in pre-lockdown risks, with median risks of 0.56 and 2.05 for lockdowns 1 and 3, respectively. In contrast, lockdown 2 did not lead to a meaningful reduction in mortality risk, likely because it was introduced earlier than the other lockdowns (risks did not increase in the first weeks of lockdown 2) or because it was in place for only four weeks. For lockdowns 1 and 3, mortality risks only stopped increasing after three weeks of lockdown, likely because there is a 21 to 25-day lag between disease onset and mortality (ONS, 2021a). The meta-analysis by Khalili et al. (2020) agrees, confirming this study's findings with an estimated 21.61 days between infection and death. While Al-Zoughool et al. (2022) have only studied the relationship between the duration and timing of lockdown with hospital cases and infection rates rather than mortality risk, their results confirm that a lockdown of fewer than 30 days may not lead to substantial reductions, which may

explain the results for lockdown 2. Furthermore, Mellone et al. (2021) also suggested a longer lockdown of about 50 days as most beneficial in reducing the number of infections and deaths in Israel and Germany.

The clustering analysis of temporal trends at a local authority level (see Section 3.4.3) in lockdown 1 revealed an urban/rural risk divide, with urban areas tending to have a higher peak risk. Hamidi et al. (2020) partially corroborate these findings by showing higher risks for larger metropolitan areas in the USA but lower risks for more densely populated counties. Hence, the urban/rural divide might be driven by multiple factors such as population density, hospital occupancy rates, and mobility within and between districts. Additionally, the geographical analysis suggested that more deprived areas tend to be at higher risk. Several studies have shown an association between higher deprivation and increased mortality risks, including Williamson et al. (2020) and Sartorius et al. (2021).

For lockdown 3, the cluster analysis indicated higher peak risks in southeast England, confirmed by the regional averages presented in Figure 3.6. The heavy concentration of elevated risks in the southeast suggests a link between the Alpha variant detected in the Kent area (South East) in September 2020 and higher levels of risk. Grint et al. (2021) adds weight to this hypothesis by finding both an increased spreading of the Alpha variant in the southeast and higher mortality risks associated with the Alpha variant when compared to earlier versions of COVID-19.

Lastly, the comparison of temporal risk trends for the three lockdowns at a national level (see Section 3.4.1) showed that the median estimated risk from the week before the introduction of lockdown 3 was much higher than that of lockdowns 1 and 2. The higher median risk could be due to a delayed implementation of lockdown 3 after the Christmas holidays, followed by more hospital patients in the first weeks of 2021 than during the peak of April 2020 (Prime Minister's Office, 2021b). Although mortality risks were at a similar level in weeks 7 through 12 after the implementation of the lockdown, the first lockdown was lifted in its seventh week, while lockdown 3 was only lifted in its twelfth week. Lockdown 3 was kept in place because the number of patients in hospitals was high (Prime Minister's Office, 2021a) despite the low mortality risks, which suggests that the relationship between mortality risks and the

number of patients in hospitals had changed from lockdown 1 to lockdown 3. Indeed, Gray et al. (2021) showed that the proportion of mortalities amongst hospitalised COVID-19 patients had decreased substantially from the first wave (lockdown 1) to the second wave (lockdowns 2 and 3). Hence, studying the impact of lockdowns on hospital admissions could be a valuable extension to this study to see if lockdowns can help relieve the stressed healthcare sector in the peak times of the pandemic.

When communicating these results, it is important to state that this study is exploratory. The study could not assess a 'causal effect' of lockdown on mortality risk as the counterfactual event of what would have happened without the lockdown could not be observed. Hence, the study comments on trends alone. Although only lockdowns 1 and 3 lasted long enough to show a reduction in mortality risks, the temporal trends at a national level were quite similar for these lockdowns, which suggests that lockdowns might have a real effect in reducing mortality risks. Since no explanatory variables were included in the model, high risks might be explained by unmeasured risk factors such as the number of care homes or levels of deprivation in each LAD. Hence, possible future work would be to examine the effects of these other covariate factors. However, if explanatory variables were to be included in the model, a great variety of them should be considered to reduce the chances of failing to account for confounding variables that could lead to incorrect estimates for the variables included in the model.

The study could be extended by considering local lockdowns or the tiered restrictions implemented in October and December of 2020 to study additional nuances of how restrictions could affect mortality risks. Alternatively, data from different countries could be considered. However, for the study of local lockdowns or lockdowns from multiple countries, the data would have to capture differences in lockdown measures. Also, when considering data from multiple countries, not only differences in institutional arrangements but also national cultural orientation should be considered (Yan et al., 2020), as the citizen's obedience to restrictions could impact the effectiveness of the lockdown. However, even when obedience is considered, the results might not be clear since people who obey the restrictions are not always guaranteed not to get infected with COVID-19. Similarly, those who disregard the restrictions

are not guaranteed to get infected.

Based on this discussion, an exciting prospect for future studies could be to link mortality risks to the number of hospitalisations or infections when studying the impact of lockdown. However, the number of infections should be considered carefully as many infections will not have been reported. Additionally, information on vaccination status and the particular variant of COVID-19 that led to death could be considered in the analysis of temporal trends in mortality risks following the introduction of the lockdown, depending on data availability.

National lockdowns were implemented as a non-pharmaceutical intervention to reduce the spreading of the virus. In the summer of 2021, COVID-19 vaccines became available at large scale in the UK. Hence, as a natural continuation of studying public health aspects of the COVID-19 pandemic in the UK, Chapter 4 presents an analysis of a COVID-19 vaccine-related topic. Specifically, the study analyses COVID-19 vaccine fatigue in Scotland, identifying trends in attrition rates for the second and third doses and identifying differences by age, sex, and council area.

Chapter 4

COVID-19 vaccine fatigue in Scotland: How do the trends in attrition rates for the second and third doses differ by age, sex, and council area?

4.1 Introduction

The introduction of Chapter 3 discussed the COVID-19 pandemic and its devastating impact on physical health, resulting in more than 6 million deaths worldwide (Johns Hopkins University & Medicine, 2022). In the first year of the pandemic, no effective medical treatment was available for the novel coronavirus disease, forcing governments to implement non-pharmaceutical interventions, such as lockdowns (see Chapter 3), to limit its spread. Pharmaceutical companies developed COVID-19 vaccines at an unprecedented speed, allowing for the start of large-scale vaccination programmes by the end of 2020 to slow down the spread of the virus, reduce the severity of disease progression in those vaccinated who become infected, and ultimately reach herd immunity. The vaccine developed by Pfizer/BioNTech was the first to be granted

emergency use by UK regulators on 2nd December 2020, making the UK the first country to approve a COVID-19 vaccine that was tested in a large clinical trial (Ledford et al., 2020). By June 2022, 150 million doses of COVID-19 vaccine had been administered in the UK (UK Health Security Agency, 2023a), and more than 11.4 billion doses had been administered worldwide (Johns Hopkins University & Medicine, 2022).

As of 27th May 2022, there are 160 vaccines in clinical development, with 11 vaccines in phase IV (WHO, 2022), meaning that they have been approved and licensed (CDC, 2014). COVID-19 vaccines are either genetic (i.e., DNA- or RNA-based), or use live-attenuated viruses, inactivated viruses, protein subunits, or replication-deficient vectors as platforms (Nagy and Alhatlani, 2021). In the time frame of this study, Scotland (the study area) had authorised only the vaccines developed by Pfizer/BioNTech, Oxford/AstraZeneca, and Moderna for use (PHS, 2022). These vaccines started to roll out on 8th December 2020, 4th January 2021, and 7th April 2021, respectively (The Scottish Parliament, 2022). The vaccines developed by Pfizer/BioNTech and Moderna are RNA-based, while that of Oxford/AstraZeneca uses a non-replicating viral vector as a platform (Ndwandwe and Wiysonge, 2021). All three vaccines have completed phases I-III of their clinical trials. The respective developers have published their phase III efficacy results, and continued phase IV studies after the vaccines had been approved and licensed (WHO, 2022). Phase III of the respective clinical trials resulted in an estimated efficacy of 95% for the Pfizer/BioNTech vaccine, 94% for the Moderna vaccine, and 62-90% for the Oxford/AstraZeneca vaccine after the administration of two doses (Costanzo et al., 2022).

While Phase III of the clinical trials promised high efficacy of the vaccines against COVID-19, vaccine efficacy and the duration of protection might vary for the different virus variants. Bernal et al. (2021) compared the effectiveness of the Pfizer/BioNTech and Oxford/AstraZeneca vaccines against the Alpha and Delta variants of the COVID-19 virus. For both vaccines, the estimated vaccine effectiveness was greater than 60% for both variants, with the Pfizer/BioNTech vaccine having overall higher effectiveness than the Oxford/AstraZeneca vaccine. Furthermore, their results suggested that

both vaccines were much more effective after administering a second dose than receiving only the first dose. Regarding the waning of vaccine effectiveness, Andrews et al. (2022b) suggested that the protection against COVID-19-related hospitalisation and death remained high 20 weeks after administering two doses of either the Pfizer/BioNTech or Oxford/AstraZeneca vaccine. Still, third doses (commonly referred to as *booster* vaccines) were shown to provide increased protection not only against symptomatic COVID-19 infection, but also against hospitalisation and death (Andrews et al., 2022a). Thus, receiving additional doses of the vaccine appears to be beneficial in the fight against the virus.

According to the UK government website (https://coronavirus.data.gov.uk/details/vaccinations), by 17th May 2022, 4.51 million people in Scotland had received the first dose of vaccine, and 4.20 million people were fully vaccinated, meaning that they had received a first and second dose. Thus, roughly 300,000 people received the first but not the second dose. Furthermore, only 3.53 million people received the third dose (booster). Note that three months after receiving the second dose of the COVID-19 vaccine, everyone over 16 becomes eligible for such a booster vaccine (NHS, 2022). Hence, with 4.01 million people in Scotland receiving the first and second doses before the end of 2021, almost 500,000 people of those eligible to take up the booster vaccine by 17th May 2022 did not do so.

The substantial number of people who opted not to receive either their second or third doses of the vaccine suggests they have experienced something similar to *vaccine hesitancy*, which MacDonald (2015) defines as a "delay in acceptance or refusal of vaccination despite availability of vaccination services". The author acknowledges that vaccine hesitancy is complex and case-specific, yet claims three factors influence vaccine hesitancy generally: confidence, complacency, and convenience. For COVID-19 specifically, vaccine confidence could be affected by the vaccines' rapid development, unforeseen future side effects, and concerns about vaccines being promoted for pharmaceutical gains, which Kumari et al. (2021) have identified as factors likely related to COVID-19 vaccine hesitancy. Kourlaba et al. (2021) suggested that the source of information impacted vaccine hesitancy, saying that those "informed by mass media and official national and state websites were more likely to be willing

to get vaccinated over those informed by social media, internet, or other sources". Mewhirter et al. (2022) identified a lack of trust and knowledge about the COVID-19 vaccines as one of the key factors associated with vaccine hesitancy. Additionally, they claimed that seeing a COVID-19 infection as a significant risk to one's health or the health of people in one's social network might be linked to higher vaccine acceptance, which explains how complacency could play a role in COVID-19 vaccine hesitancy. Lastly, the findings by Mayer et al. (2022) suggest that convenience could affect vaccine hesitancy, as they found that having to take time off work or travel to get the vaccine was associated with higher vaccine hesitancy.

Several studies suggested associations between vaccine hesitancy and demographic variables. For example, people of older age (Mewhirter et al., 2022), higher socioeconomic status (Kumari et al., 2021), and higher levels of education (Liu and Li, 2021) appeared less hesitant to get vaccinated. Lee and Huang (2022) found that areas with higher population densities appeared to have higher vaccination rates, which is supported by Khubchandani et al. (2021) who claimed that rural populations were more hesitant. Lastly, the review by Troiano and Nardi (2021) showed that several studies indicated women tended to show greater vaccine hesitancy than men.

In this study, only the population who received at least a first dose of vaccine is considered, and hence their delay in acceptance or refusal of subsequent doses is referred to as *vaccine fatigue*, as a distinction from vaccine hesitancy which refers to the acceptance or refusal of any doses. A discussion on similarities and differences between the previously mentioned associations of demographic variables with vaccine hesitancy and the associations between vaccine fatigue and the variables sex, transition between doses, age, and area examined in this study are presented in Section 4.5.

It appears that vaccine fatigue has so far only been considered in the study of booster vaccine uptake. For example, Rzymski et al. (2021) and Sønderskov et al. (2021) suggested that older people had a higher willingness to receive a booster vaccine in Poland and Denmark, respectively. While Rzymski et al. (2021) found that females in Poland were associated with higher willingness than males to receive the booster vaccine, Gaffney et al. (2022) showed that in the USA, females had lower odds of booster receipt than males. Rzymski et al. (2021) and Dziedzic et al.

(2022) identified a reduced willingness to receive the booster vaccine in the Polish population previously infected with the virus, and Geers et al. (2022) found that in the USA, "booster intentions were not associated with the number of side effects, side effect intensity, or the occurrence of an intense side effect from a primary COVID-19 vaccine". Paul and Fancourt (2022) suggested that in the UK, having a lower educational qualification, lower socio-economic status, and age below 45 years were associated with greater uncertainty and unwillingness to receive the booster vaccine, while their study suggested no such association with sex.

The novelty of this study is that in the examination of vaccine fatigue, it distinguishes between the proportions of people who received the first dose of the vaccine but not a second dose, and those who received the first and second doses but not the third (booster). Throughout this chapter, the proportions of people leaving the vaccination programme are referred to as *attrition rates*, and mathematical notation is provided in Section 4.2.3. The study analyses the impact of the demographic variables age and sex on the attrition rates within Scotland while accounting for spatial variation across council areas and allowing trends to differ in the two transitions (from doses 1 to 2 and 2 to 3). The following are the main questions of interest:

1. Are there any trends in attrition rates by age group, and does this vary by sex?

2. Are there any spatial patterns in attrition rates by council area, and does this vary by sex?

3. How do these trends and patterns compare across the two transitions (from doses 1 to 2 and 2 to 3)?

The motivating questions are of particular interest to health authorities, as the detected trends and patterns in attrition rates can help vaccination campaigns target those demographic groups that are most likely to drop out of the vaccination programme. The targeted campaigns could improve the effort to make a larger proportion of the population continue receiving doses of the COVID-19 vaccine to reduce the spreading of the virus to protect the most vulnerable and ultimately reach herd immunity. Moreover, estimating the trends in attrition rates allows to identify sub-

groups that do not follow the overall trends which might help investigate the possible reasons for why in some subgroups the uptake was lower than expected.

The remainder of this chapter is organised as follows. Section 4.2 defines attrition rates and presents the data from the vaccination programme in Scotland. Section 4.3 describes the methodology that is applied, while Section 4.4 presents the results of the study. Section 4.5 provides a discussion of the study results, limitations, and possible future work.

4.2 Motivating study

4.2.1 Study region

The region of this study is Scotland, UK, which is partitioned into 32 council areas. According to mid-2020 population estimates, the population sizes of these council areas range from 22,400 people on the Orkney Islands to more than 635,000 people in Glasgow City, and the sizes of the geographical areas range from 60 square kilometres in Dundee City to 25,653 square kilometres in the Highlands (ONS, 2021c).

4.2.2 Vaccination data

Public Health Scotland (PHS) provides data on the daily numbers of vaccinations for each dose (1st, 2nd, 3rd) of vaccine administered by sex (female, male), age group (18 to 29, 30 to 39, 40 to 49, 50 to 54, 55 to 59, 60 to 64, 65 to 69, 70 to 74, 75 to 79, 80+), and council area (32 local authorities), which are publicly available at `https://www.opendata.nhs.scot/dataset/covid-19-vaccination-in-scotland`. These data are available from 8th December 2020, when the Scottish vaccination programme began, and the data used in this study end on 19th April 2022. For this study only the cumulative counts of vaccine uptake in adults at 19th April 2022 are used, since vaccines became gradually available based on age, vulnerability, or exposure to the virus (Department of Health & Social Care, 2021b), making the analysis of temporal trends in vaccine uptake impractical. Specifically, the data relate to the number of people who received a first, first and second, or first, second, and

third (booster) dose by the cutoff date. By 19th April 2022, 4,013,535 (∼90%) people of age 18 and over had received the first dose, 3,879,067 (∼87%) had received the first and second doses, and 3,397,185 (∼77%) had received the first, second, and third doses of the vaccine (the proportions are based on a mid-2020 population estimate of 4,439,078 people aged 18 and over in Scotland, NRS, 2021). Note that the data do not specify the type of vaccine administered. However, at the time of this study, only the vaccines developed by Pfizer/BioNTech, Oxford/AstraZeneca, and Moderna had been offered in Scotland (PHS, 2022).

In this study, the population under consideration consists of only those individuals in Scotland who had received at least a first dose of vaccine. The youngest age group in this study, 18 to 29 years old, was the last age group who were able to book booster appointments, starting 15th December 2021 (Scottish Government, 2021b). Thus, by 19th April 2022, everyone in Scotland aged 18 years and older should have had enough time to receive all three doses if desired. Hence, it is likely that most people who did not continue to receive their second or third doses of vaccine experienced *vaccine fatigue* (delay or refusal to accept another dose of the vaccine), rather than being unable to take up the vaccination due to health restrictions or limitations in the number of vaccines being offered.

4.2.3 Attrition Rates

Let Y_{sgkd} denote the cumulative number of people of sex s ($s = \text{female}, \text{male}$) in age group g ($g = 1, 2, \ldots, G(= 10)$) from council area k ($k = 1, 2, \ldots, K(= 32)$) who have received a d^{th} ($d = 1, 2, 3$) dose of the COVID-19 vaccine by 19th April 2022. This study focuses on *attrition rates*, which refer to the proportion of people who leave the vaccination programme after receiving a preceding dose. The doses of the COVID-19 vaccine form a sequence with states 'dose 1', 'dose 2', and 'dose 3'. Adopting a commonly used notation from sequence analysis (Liao et al., 2022), this study uses the term *transition* to denote a change of state (from doses 1 to 2 or 2 to 3) in this sequence. The attrition rate in the transition from dose $(d-1)$ to dose d of sex s,

age group g, and council area k is denoted by $A_{sgk}^{(d-1)d}$ and computed as

$$A_{sgk}^{(d-1)d} = 1 - \frac{Y_{sgkd}}{Y_{sgk(d-1)}}, \text{ for } d = 2, 3. \tag{4.1}$$

Note that the exploratory analysis that follows will present the attrition rates as percentages to give a more intuitive interpretation, thus multiplying Equation (4.1) by 100. The focus of this study is on the attrition rates between doses 1 and 2 (A_{sgk}^{12}), and between doses 2 and 3 (A_{sdk}^{23}), to see if there are any trends over the different age groups or patterns across the council areas, and if this varies by sex or across the two transitions. Note that the observed attrition rates are likely to be noisy estimates of the true underlying attrition rates, since the observed vaccination counts $Y_{sgk(d-1)}$ and Y_{sgkd} may be subject to data processing errors or random variation. Possible associations between the so observed attrition rates and the other variables age group, sex, transition between doses, and council area are explored descriptively in Section 4.2.4. The models presented in Section 4.3 estimate the shape and magnitude of these associations to obtain smoothed estimates of the expected attrition rates.

4.2.4 Associations between attrition rates and other variables

The goal of this study is to obtain smooth estimates of the expected attrition rates that reflect the associations between the observed attrition rates and the variables age group, sex, transition between doses, and council area. This section provides a descriptive exploratory analysis to identify these associations. Note that other variables could be used to obtain smoothed estimates, however the PHS data presented in Section 4.2.2 only provides the vaccination counts by age group, sex, dose, and council area, and therefore only these variables are considered at this stage of the analysis.

4.2.4.1 How do the attrition rates vary by age?

Figure 4.1 presents transition-specific boxplots that show the observed attrition rates over the 32 council areas for each combination of age group and sex. The plots

show that attrition rates tend to decrease with increasing age and are generally much higher in the transition from doses 2 to 3 than in the transition from doses 1 to 2. In both transitions, the sub-group with the highest observed attrition rates is that of males aged 18 to 29 years, with median attrition rates of 8.84% (from doses 1 to 2) and 32.75% (from doses 2 to 3).

Attrition rates tend to be higher for males than females in age groups 54 years and younger, and this trend reverses for age groups 75 years and older, where males tend to have lower attrition rates than females. This is not easy to see in Figure 4.1, which uses the same scale for all age groups; Section B.1 of the appendix provides additional plots that show the reversing trend more clearly. Furthermore, the differences in attrition rates between the two sexes tend to be larger for younger age groups, and they become more similar with increasing age, which suggests a possible interaction effect between age group and sex on the attrition rates. For age groups 59 years and younger, the variation within (i.e., over council areas) and between age groups is much larger in the transition from doses 2 to 3 than in the transition from doses 1 to 2, suggesting a possible interaction effect between age group and transition. Similarly, the differences between sexes tend to be much larger for age groups 54 years and younger in the transition from doses 2 to 3 than in the transition from doses 1 to 2, suggesting a possible interaction between sex and transition in the age group effects.

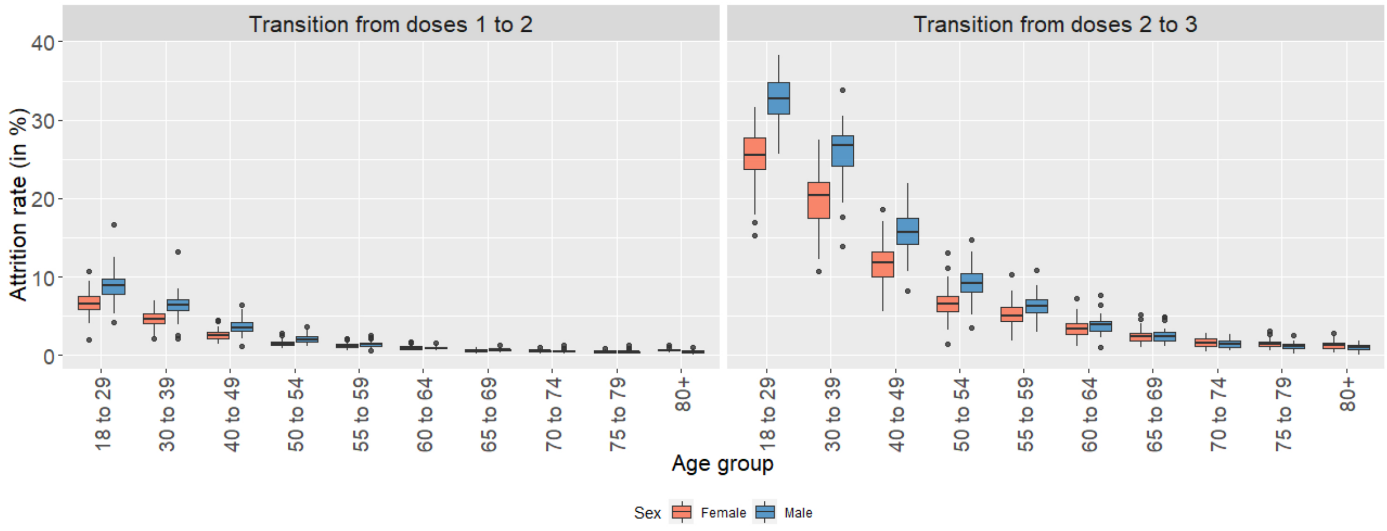


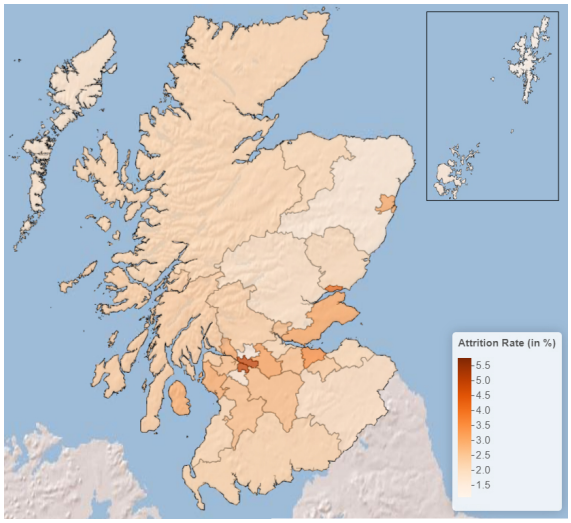
Figure 4.1: Boxplots of the attrition rates (in %) for the different age groups, by sex and transition.

4.2.4.2 How do the attrition rates vary by council area?

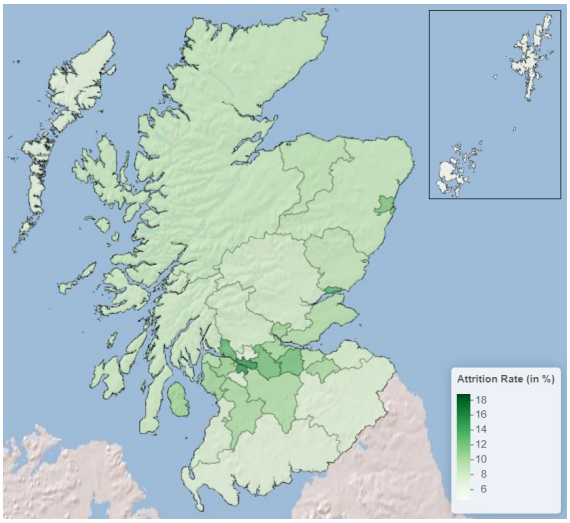
Next, consider the differences in attrition rates by council area and sex for the transitions from doses 1 to 2 and 2 to 3. Averaging over age group, the number of people of sex s in council area k who received up to the d-th dose of vaccine is computed by $Y_{skd} = \sum_{g=1}^{10} Y_{sgkd}$, yielding an attrition rate from dose $(d-1)$ to dose d of

$$A_{sk}^{(d-1)d} = 1 - \frac{Y_{skd}}{Y_{sk(d-1)}}, \text{ for } d = 2, 3. \tag{4.2}$$

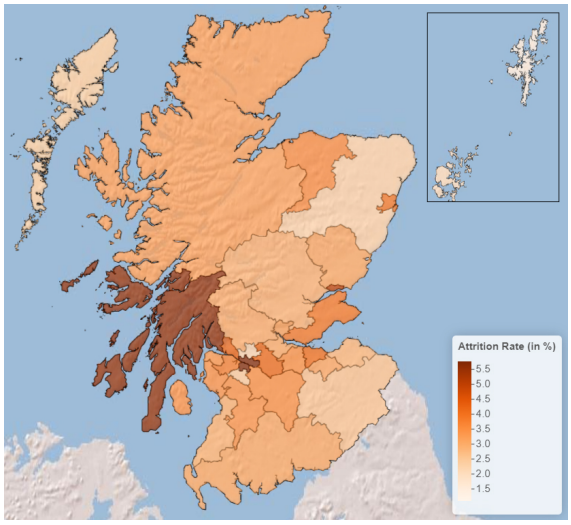
(A) Female, doses 1 to 2.



(B) Female, doses 2 to 3.



(C) Male, doses 1 to 2.



(D) Male, doses 2 to 3.

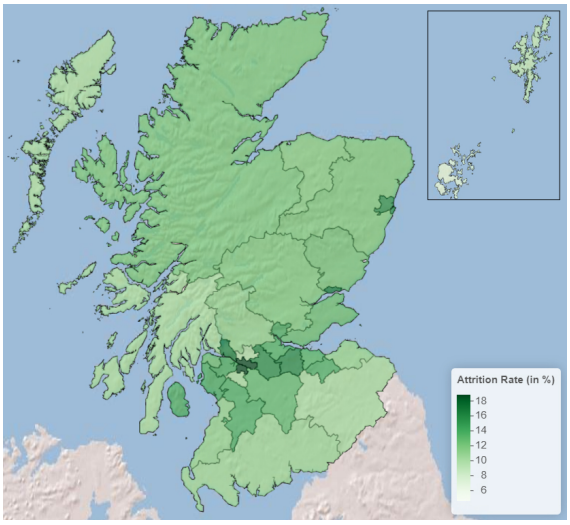


Figure 4.2: Maps showing the attrition rates (in %) by council area, sex, and transition.

Figure 4.2 presents the spatial patterns in the attrition rates (in %) across council areas for each combination of sex and transition (from doses 1 to 2 and 2 to 3). Note

that the Orkney Islands and Shetland Islands are downscaled and presented in a box rather than at their true location in the North of Scotland to make the maps more compact. The full map of Scotland is included in Section B.2 of the appendix, for reference. The attrition rates range from 1.25% to 4.90% for females and from 1.46% to 5.87% for males in the transition from doses 1 to 2. In the transition from doses 2 to 3, the attrition rates range from 5.20% to 16.00% for females and from 7.72% to 19.28% for males. Note that the colour scales in Figure 4.2 are the same for females and males in each transition, as the ranges of attrition rates are relatively similar for the two sexes. However, since the ranges of attrition rates are quite different in the two transitions, the colour scale for the transition from doses 1 to 2 differs from that of the transition from doses 2 to 3. Thus, while the colour distributions allow for a direct comparison of the attrition rates for the two sexes in each transition, they can only be used for a relative comparison between transitions.

In each transition, the highest attrition rates were observed for males in Glasgow City, at 5.87% (from doses 1 to 2) and 19.28% (from doses 2 to 3), respectively. These attrition rates imply that in Glasgow City, roughly one in 20 males who received the first dose did not receive the second dose, and approximately one in five males who received the second dose did not receive the third. The pattern of shading on the maps look relatively similar for each sex when moving from the first transition to the second, suggesting that there could be consistent spatial patterns in the attrition rates across the two transitions. However, for males, there are a few council areas whose attrition rates changed drastically from the first transition to the second, compared to the changes in most other council areas. For example, males in Argyll and Bute had one of the highest attrition rates (5.69%) in the transition from doses 1 to 2 but one of the lower attrition rates (10.65%) in the transition from doses 2 to 3, suggesting that relative to the other council areas, more people left the vaccination programme after their first dose but fewer left after their second. Such extreme differences in a few council areas suggest that there might be an interaction effect between council area and transition on attrition rates. While attrition rates tend to be higher for males than females, the spatial distributions of attrition rates look relatively similar for the two sexes (except for Argyll and Bute in the first transition). Hence, there

may not be much of an interaction between the variables council area and sex on attrition rates.

A Moran's I test (recall Section 2.5.1.1) can be used as a formal check of whether there is any spatial autocorrelation in the attrition rates. Here, Moran's I test statistic for sex s and the transition from dose $(d-1)$ to dose d is computed as

$$I_{sd} = \frac{K \sum_{k=1}^{K} \sum_{j=1}^{K} w_{kj} \left(A_{sk}^{(d-1)d} - \bar{A}_{s}^{(d-1)d} \right) \left(A_{sj}^{(d-1)d} - \bar{A}_{s}^{(d-1)d} \right)}{\left(\sum_{k=1}^{K} \sum_{j=1}^{K} w_{kj} \right) \sum_{k=1}^{K} \left(A_{sk}^{(d-1)d} - \bar{A}_{s}^{(d-1)d} \right)^2}, \tag{4.3}$$

where $\bar{A}_{s}^{(d-1)d} = \frac{1}{K} \sum_{k=1}^{K} A_{sk}^{(d-1)d}$ is the average observed attrition rate over all council areas, for sex s and the transition from doses $(d-1)$ to d, and w_{kj} indicates whether or not council areas k and j share a border; i.e., $w_{kj} = 1$ if council areas k and j share a border, and $w_{kj} = 0$, otherwise. The Moran's I statistic can only be computed for data where each area has at least one weight different from zero; i.e., for each area k, there is at least one area $j \neq k$ such that $w_{kj} > 0$. Since only council areas in mainland Scotland share borders with other council areas, the Orkney Islands, Shetland Islands, and the Western Isles Na h-Eileanan Siar (the Outer Hebrides) are linked to their nearest council areas before computing the test statistic. This artificially links the Shetland Islands and the Orkney Islands, the Orkney Islands and the Highlands, and the Western Isles Na h-Eileanan Siar and the Highlands. The computed test statistics are -0.0576 (female, from doses 1 to 2), -0.0176 (male, from doses 1 to 2), 0.0888 (female, from doses 2 to 3), and 0.0361 (male, from doses 2 to 3).

To test if the spatial autocorrelations in each combination of sex and transition are statistically significant, a permutation test is performed based on 10,000 random permutations of the data, using Monte Carlo simulation. That is, in each of the 10,000 iterations the observed attrition rates are randomly assigned to the council areas, and Moran's I statistic is computed for each permuted set. This distribution generated under independence is then compared to the original statistic that was computed for the observed data to yield a p-value against independence. For each combination of sex and transition, the permutation test is evaluated using a significance level of $\alpha = 0.05$. The p-values for the test statistics are 0.5746 (female, from doses 1 to 2),

0.4250 (male, from doses 1 to 2), 0.1426 (female, from doses 2 to 3), and 0.2653 (male, from doses 2 to 3). Since all of these p-values are much larger than 0.05, none of the permutation tests suggests there is statistically significant spatial autocorrelation in the observed attrition rates.

However, this lack of significance is possibly due to the very small number of council areas (only 32), meaning that significant effects are harder to find. Hence, there remains a possibility that there might be some spatial structure in the data, and it is desirable to consider a flexible model for this study that estimates the strength of the spatial correlation from the data.

4.3 Methodology

The model used in this study is designed to identify the trends and patterns in attrition rates by age group and council area, and to examine how these differ by sex and transition (from doses 1 to 2 and 2 to 3). In describing the proposed model, Section 4.3.1 describes the data likelihood and Section 4.3.2 presents the structure of the random effects. Section 4.3.3 discusses prior specifications and Section 4.3.4 presents simplified versions of the random effects structures that are considered to prevent the model from overfitting.

4.3.1 Data Likelihood

The goal is to estimate the attrition rates from doses 1 to 2 and doses 2 to 3, so for each sex s, age group g, and council area k, I simultaneously model the number of people who received dose 2 and dose 3, $\{Y_{sgk2}, Y_{sgk3}\}$, given the number of people who received the first dose of the COVID-19 vaccine $\{Y_{sgk1}\}$. I propose the following joint data likelihood model.

$$\begin{aligned} f(\{Y_{sgk2}, Y_{sgk3}\} \mid \{Y_{sgk1}\}) &= \prod_s \prod_g \prod_k f(Y_{sgk2}, Y_{sgk3} \mid Y_{sgk1}) \\ &= \prod_s \prod_g \prod_k f(Y_{sgk3} \mid Y_{sgk2}) \times f(Y_{sgk2} \mid Y_{sgk1}), \end{aligned} \tag{4.4}$$

where the second equality follows from the conditional probability formula and the assumption that the number of people who receive a third dose of the vaccine only depends on the number of people who received a second dose but not on the number of people who received a first dose. The model assumes independence between the two transitions for age, sex, and council area, conditional on the parameter values of the function $f(\cdot)$. The dependencies in the parameter values are accounted for by autocorrelated random effects as presented below.

Assume that a randomly selected individual from sex s, age group g, and council area k has a probability of $1 - \alpha_{sgk}^{(d-1)d}$ to receive dose d of the vaccine, given they have received dose $(d-1)$. Furthermore, the number of people who receive dose d of the vaccine has to be less than or equal to the number of people who have received dose $(d-1)$. That is, $Y_{sgkd} \leq Y_{sgk(d-1)}$ $(d = 2, 3)$ for each sex s, age group g, and council area k. Hence, given the number of people who have received dose $(d-1)$ of vaccine and the probability $1 - \alpha_{sgk}^{(d-1)d}$ that these individuals receive dose d, the number of people who receive dose d can be modelled using a binomial distribution.

Motivated by the three leading questions, Section 4.2.4 identified possible relationships between the attrition rates and age group, sex, council area, and transition between doses, as well as possible interaction effects between age group and sex, age group and transition, council area and sex, and council area and transition on attrition rates. Hence, in the most general model considered, the logit transformation of $1 - \alpha_{sgk}^{(d-1)d}$ is modelled via a linear model that contains effects for the before-mentioned variables and interaction terms. Therefore, the most general form of the model I propose is given by

$$Y_{sgkd} \mid Y_{sgk(d-1)} \sim \text{Binomial}\Big(Y_{sgk(d-1)}, 1 - \alpha_{sgk}^{(d-1)d}\Big), \tag{4.5}$$

$$\log\left(\frac{1 - \alpha_{sgk}^{(d-1)d}}{\alpha_{sgk}^{(d-1)d}}\right) = \beta_0 + \gamma_s + \psi_d + \delta_g^{(sd)} + \phi_k^{(sd)},$$

where $Y_{sgk(d-1)}$ denotes the number of binomial trials, i.e. the number of people who received dose $(d-1)$ and are therefore eligible to receive dose d. Here, $\alpha_{sgk}^{(d-1)d}$ denotes the probability that a randomly selected individual of sex s and age group g in council

area k quits the vaccination programme after dose $(d-1)$, and it is easy to show that $\mathbf{E}\left[A_{sgk}^{(d-1)d} \mid Y_{sgk(d-1)}\right] = \alpha_{sgk}^{(d-1)d}$ (the derivation is provided in Section B.3 of the appendix). Thus, the observed attrition rate $A_{sgk}^{(d-1)d}$ is an unbiased estimator of the true underlying attrition rate $\alpha_{sgk}^{(d-1)d}$, which is modelled via a generalised linear model by taking the logit transformation of $1-\alpha_{sgk}^{(d-1)d}$. The parameters in the generalised linear model of Equation (4.5) are an intercept term β_0 for females in the transition from doses 1 to 2, a fixed sex effect γ_s of being male as opposed to being female (i.e., $\gamma_s = 0$ for $s =$ female), a fixed transition effect ψ_d of being in the transition from doses 2 to 3 as opposed to the transition from doses 1 to 2 (i.e., $\psi_d = 0$ for $d = 2$), a random age group effect $\delta_g^{(sd)}$ that is allowed to vary by sex and transition, and a random council area effect $\phi_k^{(sd)}$ that is allowed to vary by sex and transition. These parameters are included to answer the three motivating questions. That is, the estimated effects will suggest whether the attrition rates differ by age group or council area, if there are differences in these trends and patterns for the two sexes, and if any of these dynamics differ for the two transitions.

The odds in favour of attrition to non-attrition can be obtained from Equation (4.5) as

$$\begin{aligned}\alpha_{sgk}^{(d-1)d} \Big/ \left(1-\alpha_{sgk}^{(d-1)d}\right) &= \exp\Big(-\Big(\beta_0+\gamma_s+\psi_d+\delta_g^{(sd)}+\phi_k^{(sd)}\Big)\Big) \\ &= \exp(-\beta_0)\times\exp(-\gamma_s)\times\exp(-\psi_d) \\ &\quad\times\exp\big(-\delta_g^{(sd)}\big)\times\exp\Big(-\phi_k^{(sd)}\Big).\end{aligned} \tag{4.6}$$

For each combination of sex and transition, sum-to-zero constraints are placed on the random age group ($\delta_g^{(sd)}$) and council area ($\phi_k^{(sd)}$) effects to prevent them from confounding with the intercept term β_0 and the fixed effects γ_s and ψ_d. Thereby, for a fixed transition from doses $(d-1)$ to d, the odds ratio of attrition if one is male as opposed to female, averaged over all age groups and council areas, reduces to $\exp(-\gamma_s)$ (see Section B.4 in the appendix for details). Hence, for a fixed transition from doses $(d-1)$ to d, the term $\exp(-\gamma_s)$ can be interpreted as the average odds ratio of attrition if one is male compared to female. Similarly, for a fixed sex s, the term $\exp(-\psi_d)$ can be interpreted as the average odds ratio of attrition if one is in

the transition from doses 2 to 3 compared to the transition from doses 1 to 2.

The random effects are included to capture the differences in the odds in favour of attrition by age group and council area, allowing for the fact that neighbouring 'groups' may be correlated. For a fixed council area k, sex s, and transition from doses $(d-1)$ to d, the odds ratio of attrition for age group g compared to the sex- and transition-specific average odds across all age groups is $\exp\left(-\delta_g^{(sd)}\right)$, and in the most general model this factor is allowed to differ for the two sexes and transitions. For example, an odds ratio of $\exp\left(-\delta_g^{(sd)}\right) = 1.2$ suggests that for sex s and the transition from doses $(d-1)$ to d, the odds in favour of attrition in age group g are 1.2 times higher than the average odds across all age groups. Similarly, for a fixed age group g, sex s, and transition from dose $(d-1)$ to d, the odds ratio of attrition for council area k compared to the sex- and transition-specific Scottish average is $\exp\left(-\phi_k^{(sd)}\right)$. Section B.4 in the appendix provides mathematical derivations of the interpretations of the fixed and random effects in the model. To clarify, since the model estimates all fixed and random effects simultaneously, the resulting odds ratios are affected by the terms included in the model. From here on in, I consider it implicit that the odds ratios are, in that respect, conditional on all variables in the model.

4.3.2 Structure of the random age group and council area effects

The exploratory analysis in Section 4.2.4 showed that the observed attrition rates $\left\{A_{sgk}^{(d-1)d}\right\}$ decrease smoothly as age increases (see Figure 4.1) and that there is a possibility that they might change somewhat smoothly across the council areas (see Figure 4.2). However, the permutation tests suggest that the spatial autocorrelation in the attrition rates is not statistically significant for any of the combinations of sex and transition. Hence, BYM2 models (recall Section 2.6.1) are proposed for estimating the correlated sex- and transition-specific random age group effects $\boldsymbol{\delta}^{(sd)} = \left(\delta_1^{(sd)}, \ldots, \delta_G^{(sd)}\right)$ and council area effects $\boldsymbol{\phi}^{(sd)} = \left(\phi_1^{(sd)}, \ldots, \phi_K^{(sd)}\right)$. These models allow for flexible estimation, as they can take on either of two base models, correlated or independent, but also allow for a mixture of the two (Simpson et al.,

2017). The BYM2 model of the random age group effects can be written as

$$\begin{aligned}
\boldsymbol{\delta}^{(sd)} &= \frac{1}{\sqrt{\tau_\delta}}\Big(\sqrt{1-\rho_\delta}\boldsymbol{v}_\delta^{(sd)} + \sqrt{\rho_\delta}\boldsymbol{u}_\delta^{*(sd)}\Big), \\
\boldsymbol{v}_{\boldsymbol{\delta}}^{(sd)} &\sim N(\mathbf{0}, \boldsymbol{I}), \\
\boldsymbol{u}_\delta^{*(sd)} &\sim N\big(\mathbf{0}, \boldsymbol{Q}^*(\boldsymbol{W})^-\big),
\end{aligned}$$

and the BYM2 model of the random council area effects is similarly expressed as

$$\begin{aligned}
\boldsymbol{\phi}^{(sd)} &= \frac{1}{\sqrt{\tau_\phi}}\Big(\sqrt{1-\rho_\phi}\boldsymbol{v}_\phi^{(sd)} + \sqrt{\rho_\phi}\boldsymbol{u}_\phi^{*(sd)}\Big), \\
\boldsymbol{v}_{\boldsymbol{\phi}}^{(sd)} &\sim N(\mathbf{0}, \boldsymbol{I}), \\
\boldsymbol{u}_\phi^{*(sd)} &\sim N\big(\mathbf{0}, \boldsymbol{Q}^*(\boldsymbol{M})^-\big),
\end{aligned}$$

where $\boldsymbol{u}_\delta^{*(sd)}$ and $\boldsymbol{u}_\phi^{*(sd)}$ are vectors of scaled correlated random effects (according to the age group and spatial structures defined by the matrices $\boldsymbol{W}$ and $\boldsymbol{M}$), and $\boldsymbol{v}_\delta^{(sd)}$ and $\boldsymbol{v}_\phi^{(sd)}$ are vectors of independent random effects of lengths G and K, respectively. The precision parameters τ_δ and τ_ϕ control the overall variation in the random effects, and the mixing parameters ρ_δ and ρ_ϕ control how much of the variation is accounted for by the correlated and independent components, respectively. When the mixing parameters are close to zero, the independent components account for most of the variation in the corresponding random effects $\boldsymbol{\delta}$ and $\boldsymbol{\phi}$. As the values of the mixing parameters increase, more variation is accounted for by the correlated components, and parameter values close to one imply strong age group or spatial dependence. Mixtures between these base models (independence vs strong autocorrelation) are achieved when the mixing parameters are not zero or one.

Note that the random effects $\boldsymbol{v}_\delta^{(sd)}$ and $\boldsymbol{u}_\delta^{*(sd)}$ that make up the age group effects $\boldsymbol{\delta}^{(sd)}$ are estimated for each combination of sex and transition. However, the hyperparameters τ_δ and ρ_δ are estimated over all observations from the two sexes and transitions. Since there are only 10 age groups in the data, assuming the hyperparameters to be the same for each combination of sex and transition allows for a larger number of observations in the estimation of the hyperparameter values. The same logic applies to the estimation of the hyperparameters τ_ϕ and ρ_ϕ of the random

council area effects $\boldsymbol{\phi}^{(sd)}$, since there are only 32 council areas in the data.

The correlated random council area effects $\boldsymbol{u}_\phi$ are assigned an ICAR prior distribution with a binary adjacency matrix $\boldsymbol{M}$ (recall Equation 2.13) based on the border sharing rule. Note that each areal unit must have at least one adjacent area in this model. In order to include the Orkney Islands, the Shetland Islands, and the Western Isles Na h-Eileanan Siar in the model fitting process, they are treated as adjacent to their respective nearest council areas, artificially linking the Shetland Islands and the Orkney Islands, the Orkney Islands and the Highlands, and the Western Isles Na h-Eileanan Siar and the Highlands. A sum-to-zero constraint ($\mathbf{1}\boldsymbol{u}_\phi^T = 0$, where $\mathbf{1}$ is a vector of ones) is placed on these random effects to prevent confounding with the intercept term β_0. For the ICAR prior distribution of $\boldsymbol{u}_\phi$, the marginal variance of $\left(u_{\phi k} \mid \boldsymbol{u}_{\phi(-k)}\right)$ is proportional to $\frac{1}{\sum_{j=1}^{K} m_{jk}}$, where $\boldsymbol{u}_{\phi(-k)}$ denotes all random council area effects except for that of area k (i.e., $\boldsymbol{u}_{\phi(-k)} = (u_{\phi 1}, \ldots, u_{\phi(k-1)}, u_{\phi(k+1)}, \ldots, u_{\phi K})$), and $m_{jk} = 1$ if council areas j and k share a border and $m_{jk} = 0$, otherwise. Note that the full conditional marginal variances are inversely proportional to the number of neighbouring areas. Hence, the marginal variances depend on the neighbourhood structure used in the specific application.

The structured random age group effects $\boldsymbol{u}_\delta$ are assigned a random walk prior distribution of order 1, reflecting the observation from Figure 4.1 that attrition rates tend to decrease relatively smoothly with increasing age. The random walk of order 1 can be viewed as an ICAR prior distribution with neighbourhood matrix $\boldsymbol{W}$ that indicates whether or not two age groups are adjacent, i.e. $w_{ij} = 1$ if $|i - j| = 1$ and $w_{ij} = 0$, otherwise. A sum-to-zero constraint is also applied to the random effects $\boldsymbol{u}_\delta$, again to prevent confounding with the intercept term β_0. For the random walk of order 1, the marginal variances of $\left(u_{\delta g} \mid \boldsymbol{u}_{\delta(-g)}\right)$ depend on the number of adjacent age groups. Specifically, the full conditional marginal variances of the age groups with two adjacent age groups are half of that of the random effects for age groups 18-29 and 80+ with only one adjacent age group. Thus, the full conditional marginal variances of the age group effects depend on the age group specifications in the considered application.

The structure of the correlated random effects via $\boldsymbol{W}$ and $\boldsymbol{M}$ must be defined by the researcher, and an assessment of the sensitivity of the results to this choice is provided in Section B.8 of the appendix. As an additional point of clarification, the CAR priors used in the model imply that neighbouring data points (e.g., adjacent age groups) are partially autocorrelated whilst non-neighbouring data points are treated as conditionally (but not marginally) independent given the remaining data values. Thus, these CAR priors allow for correlation between non-neighbouring data points.

Scaling the structured random effects allows for interpretability of the hyperprior distributions of the precision parameters τ_ϕ and τ_δ, and for transferability between applications (Riebler et al., 2016), so that similar hyperpriors can be used for comparable applications with differing structure matrices. Following the methodology proposed by Sørbye and Rue (2014), the correlated random effects are standardised using the generalised variance $\sigma^2_{GV}(\boldsymbol{u})$, which is computed as the geometric mean of the marginal variances, i.e.

$$\sigma^2_{GV}(\boldsymbol{u}) = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{1}{\tau_u}[\boldsymbol{Q}^-]_{ii}\right)\right), \tag{4.7}$$

for $\boldsymbol{u} = \boldsymbol{u}_\delta, \boldsymbol{u}_\phi$ with corresponding precision matrices $\boldsymbol{Q} = \boldsymbol{Q}(\boldsymbol{W}), \boldsymbol{Q}(\boldsymbol{M})$ of dimensions $n \times n = G \times G, K \times K$. The precision matrix of the random age group effects is computed as $\boldsymbol{Q}(\boldsymbol{W}) = \text{diag}(\boldsymbol{W}\mathbf{1}) - \boldsymbol{W}$ and that of the random council area effects is computed as $\boldsymbol{Q}(\boldsymbol{M}) = \text{diag}(\boldsymbol{M}\mathbf{1}) - \boldsymbol{M}$, where $\mathbf{1}$ denotes vectors of ones of the appropriate lengths G and K, respectively. The precision matrices are scaled to $\boldsymbol{Q}^*(\boldsymbol{W})$ and $\boldsymbol{Q}^*(\boldsymbol{M})$ such that $\sigma^2_{GV}(\boldsymbol{u}^*) = 1$ for $\tau_u = 1$, for $\boldsymbol{u}^* = \boldsymbol{u}^*_\delta, \boldsymbol{u}^*_\phi$. That is, the precision matrices are scaled such that for $\tau_u = 1$, the geometric means of the marginal variances of $\boldsymbol{u}^*_\delta$ and $\boldsymbol{u}^*_\phi$ are equal to one, meaning that the deviations of the random effects from the overall averages are independent of their underlying structures.

The unstructured random effects $\boldsymbol{v_\delta}$ and $\boldsymbol{v_\phi}$ are independent and identically distributed, following multivariate normal prior distributions with mean zero and covariance matrices $\boldsymbol{I}_{G\times G}$ and $\boldsymbol{I}_{K\times K}$, respectively. Since the geometric means of the marginal variances of the unstructured effects are one, they are already standardised.

4.3.3 Prior specifications

Prior distributions need to be specified for the intercept β_0 and for the fixed effects γ_s and ψ_d. Further hyperpriors need to be specified for the parameters τ_ϕ, τ_δ, ρ_ϕ, and ρ_δ. The intercept β_0 and the fixed effects γ_s and ψ_d are assigned weakly informative normal prior distributions with mean zero and a variance of 1,000. Thus, the prior distributions of the intercept and fixed effects impose only weak constraints and let the data speak for themselves. These prior specifications seem reasonable since for the attrition rates in Scotland, no information is available for the intercept or fixed effects that would justify using informative prior distributions.

The parameters τ_ϕ, τ_δ, ρ_ϕ, and ρ_δ are assigned penalised complexity (PC) hyperprior distributions (recall Section 2.3.1.1). The precisions τ_δ and τ_ϕ are assigned type-2 Gumbel prior distributions of the form

$$\pi(\tau) = \frac{\theta}{2}\tau^{-3/2}\exp(-\theta\tau^{-1/2}), \tag{4.8}$$

for $\tau \in \{\tau_\delta, \tau_\phi\}$. Thus, the prior distribution for the precision has infinite mean and variance, reflecting the simplifying prior assumptions that there are no age group or council area effects on the attrition rates. That is, since the correlated random effects $\boldsymbol{u}^*$ and independent random effects $\boldsymbol{v}$ have means of zero, increasing the precision τ to a very large value causes the joint effect of these random effects to be close to zero. Simpson et al. (2017) suggest using these improper Gumbel prior distributions instead of commonly used Gamma priors with finite expectations, as Frühwirth-Schnatter and Wagner (2010) and Frühwirth-Schnatter and Wagner (2011) show that the latter generally cause the models to overfit the data. Choosing the value of the parameter θ can be aided by considering the probability statement $P(1/\sqrt{\tau} > U) = \alpha$, resulting in $\theta = -\log(\alpha)/U$ (Riebler et al., 2016), where U and α can be chosen by the researcher. The default values in INLA are $U = 1$ and $\alpha = 0.01$, which are used in this study, since the corresponding probability statement reflects the prior assumption that there are no substantial age group or council area effects, with a probability of 99% that the standard deviations take on values smaller than one, causing the effects to be close to their zero means unless the data suggest otherwise. A sensitivity analysis

is carried out to check that the results are consistent for varying prior assumptions, and is provided in Section B.7 of the appendix.

The mixing parameters ρ_δ and ρ_ϕ are assigned another PC prior distribution, where the base model is obtained for $\rho = 0$, i.e. the model where the independent component explains all variability of the random effects. The Kullback-Leibler divergence (KLD, e.g., recall Equation 2.7) for ρ between the base model and corresponding flexible BYM2 model is used to compute the distance scale $d(\rho)$ (recall Section 2.3.1.1 for details), which is assigned an exponential prior distribution with parameter λ. In contrast to the precision parameter τ, the PC prior for ρ is not available in closed form, but can be computed in `R-INLA` for the specific neighbourhood matrix, and more details are provided in Riebler et al. (2016). The probability statement $P(\rho < U) = \alpha$ can be used to determine a reasonable value of λ, where the values of U and α can be chosen by the researcher. The default values in `R-INLA` are $U = 0.5$ and $\alpha = 0.5$, which are used in this study, since it is unknown whether or not to expect the random effects to be correlated before viewing the data. Other settings are explored in the sensitivity analysis in Section B.7 of the appendix.

4.3.4 Model simplifications

In the most general model, the random effects $\left\{\delta_g^{(sd)}\right\}$ and $\left\{\phi_k^{(sd)}\right\}$ are allowed to differ by sex and transition. However, if the differences by sex or transition are not very pronounced, making the random effects sex- and transition-specific might overfit the data. Potential overfitting can be checked by fitting models where the random age group effects $\left\{\delta_g^{(sd)}\right\}$ are simplified, taking on the forms $\left\{\delta_g^{(s)}\right\}$ which differ by sex but not by transition, $\left\{\delta_g^{(d)}\right\}$ which differ by transition but not by sex, or $\{\delta_g\}$ which are the same for the different combinations of sex and transition in each age group. Similar simplifications can be considered for the random council area effects (written as $\left\{\phi_k^{(s)}\right\}$, $\left\{\phi_k^{(d)}\right\}$, and $\{\phi_k\}$, respectively). The types of random effects that should be used is decided based on a comparison of model performance evaluated using WAIC (recall Section 2.4.3) and DIC (recall Section 2.4.2) values.

4.3.5 Inference and software

The models are fitted using the `R-INLA` package (recall Section 2.3.3). Further details on how to model areal data in `R-INLA` can be found in Moraga (2019) or Gómez-Rubio (2020). The BYM2 models that are applied in this study were proposed by Simpson et al. (2017) and further discussed by Riebler et al. (2016), and the scaling of the structured random effects was proposed by Sørbye and Rue (2014). Section B.6 in the appendix presents a proof of principle simulation to confirm that the proposed method produces reliable parameter estimates for the type of data considered in this study.

4.4 Results

Here, the most general model and the simplifications described in Section 4.3 are fitted to the data presented in Section 4.2. Table 4.1 shows the WAIC and DIC values for the full model from Equation (4.5) and several simplifications to assess which complexities of the full model are supported by the data. The upper block shows models that include only main effects age group (`age`), sex, council area (`area`), and transition (`trn`), starting with a single main effect in the top section, two main effects in the second section, three main effects in the third section, and all four main effects in the fourth section. The effects included in each section are chosen via a forward selection method based on the WAIC and DIC values. That is, the best-fitting model with lowest WAIC and DIC values from each section is used in the one below, where an additional main effect is added to the model. Of the models in the top section, the one containing only random age group effects has the lowest WAIC and DIC values which suggest that it is the best-fitting model with only one main effect. Therefore, the models in the second section with two main effects all contain random age group effects and one additional main effect. Here, the model with random age group effects and a fixed transition effect has the lowest WAIC and DIC values. Thus, the models in the third section contain random age group effects, a fixed transition effect, and one additional main effect. Adding the random council area effects results in a slightly lower WAIC value and substantially lower DIC value

than adding a fixed sex effect. The model in the fourth section contains all four main effects and of the models with only main effects, it has the lowest WAIC and DIC values, suggesting that all main effects should be included.

Table 4.1: WAIC and DIC values for the full model and a number of simplifications.

Model: $\text{logit}\left(1-\alpha_{sgk}^{(d-1)d}\right)=\beta_0+$	WAIC	DIC
`sex`	87,182	523,164
`trn`	85,604	534,575
`age`	66,703	313,027
`area`	135,530	750,824
`age + sex`	68,856	303,727
`age + trn`	31,350	49,387
`age + area`	97,589	301,079
`age + trn + sex`	27,901	39,126
`age + trn + area`	27,895	33,653
`age + trn + area + sex`	20,812	23,146
`sex + trn + age(sex) + area`	19,915	21,630
`sex + trn + age(trn) + area`	19,987	21,859
`sex + trn + age + area(sex)`	21,993	22,970
`sex + trn + age + area(trn)`	19,927	20,933
`sex + trn + age(sex) + area(trn)`	18,938	19,410
`sex + trn + age(trn) + area(trn)`	18,878	19,646
Model 1: `sex + trn + age(sex,trn) + area(trn)`	17,931	17,917
Model 2: `sex + trn + age(sex,trn) + area(sex,trn)`	18,407	17,389

The lower block shows models that include the four main effects and additional interactions, where for example `age(sex)` denotes sex-specific age group effects. In the first four models, only one set of random effects can differ by sex or transition. Allowing the age group effects to differ by sex or transition and allowing the council area effects to differ by transition reduces the WAIC and DIC values, suggesting a better model fit than for the model with all main effects but no interactions. In contrast, allowing the council area effects to differ by sex results in a higher WAIC value than for the model with all main effects but no interactions.

The last four models allow the two sets of random effects to differ by sex, transition, or both, and allowing the age group effects to differ by sex and transition results in the lowest WAIC and DIC values. The model with sex- and transition-specific age group effects and transition-specific council area effects, labelled Model 1, has the

lowest WAIC value (17,930.90) of all models considered. The most general model, labelled Model 2, where both the age group effects and council area effects are sex- and transition-specific has the lowest DIC value (17,389.36). Therefore, there are two best-fitting models according to the WAIC and DIC values, which merely differ in whether the council area effects are sex- and transition-specific or transition-specific only. Most results are presented below only for Model 1, as the corresponding results from the slightly more complex Model 2 are very similar. Model 1 appears to fit the data appropriately, as suggested by the plot of the observed attrition rates against the fitted in Section B.5 in the appendix.

For a fixed transition, the average odds ratio of attrition if one is male compared to female ($\exp(-\gamma_s)$) has a posterior mean of 1.099, with a 95% credible interval of (1.079, 1.120). Thus, for a fixed transition and averaged over all age groups and council areas, the odds in favour of attrition are approximately 1.1 times higher for males than females. Similarly, the posterior mean of the average odds ratio of attrition if one is in the transition from dose 2 to 3 compared to the transition from dose 1 to 2 ($\exp(-\psi_d)$) is estimated as 4.071, with a 95% credible interval of (3.710, 4.473). Thus, for a fixed sex and over all age groups and council areas, the odds in favour of attrition are, on average, approximately 4.1 times higher in the transition from doses 2 to 3 than from doses 1 to 2.

Comparing the WAIC and DIC values as a measure of model fit and analysing the estimated effects from the best-fitting models allows for answering the motivating questions of this study. Section 4.4.1 addresses the first question: Are there any trends in attrition rates by age group, and does this vary by sex? Section 4.4.2 looks at the second question: Are there any spatial patterns in attrition rates by council area, and does this vary by sex? The third question is specific to the age group and council area effects, and therefore Sections 4.4.1 and 4.4.2 also provide the respective answers to this question: How do these trends and patterns compare across the two transitions?

4.4.1 Are there any trends in attrition rates by age group, and does this vary by sex or transition?

The first striking observation from Table 4.1 is that of the models containing only a single main effect, the one with random age group effects has much lower WAIC and DIC values than the other models. Therefore, of the main effects considered, age has the strongest link with attrition rates, which confirms the impression from the exploratory analysis in Section 4.2.4.

The estimated posterior mean of the mixing parameter ρ_δ for Model 1 is 1.0, with a 95% credible interval of (0.998,1.000). Hence, there is a strong positive autocorrelation in the age group effects, such that the effects of age groups that are closer to each other are more similar. The trends in the age group effects are shown in Figure 4.3, which presents posterior densities of the sex- and transition-specific odds ratios of attrition for each age group g, compared to the average odds over all age groups, i.e. $\left\{\exp\left(-\delta_g^{(sd)}\right)\right\}$. The dashed line represents an odds ratio of one. The corresponding posterior means and 95% credible intervals are presented in Table 4.2. The figure and table show that the odds ratio of attrition decreases with increasing age (i.e., older age groups are more likely to receive further vaccine doses), which can be observed for both sexes and transitions.

Another striking observation is that overall, the age group effects differ for the two sexes. Table 4.1 shows that Model 1 with sex- and transition-specific age group effects has a substantially lower WAIC value than the model with the same terms but only transition-specific age group effects (WAIC values of 17,930.90 compared to 18,878.01). The most noticeable differences between the two sexes can be observed in Figure 4.3 for age groups 59 years and younger, where most of the densities of the sex-specific odds ratios barely overlap (if at all) in each of the transitions. In these age groups the odds ratios are larger for males than females. However, this trend reverses for age groups 70 years and older, where the odds ratios are larger for females than males. This trend aligns with the observations from Section 4.2.4.1 and the additional plots in Section B.1 of the appendix, which show that the observed attrition rates are higher for females than males in age groups 70 years and older.

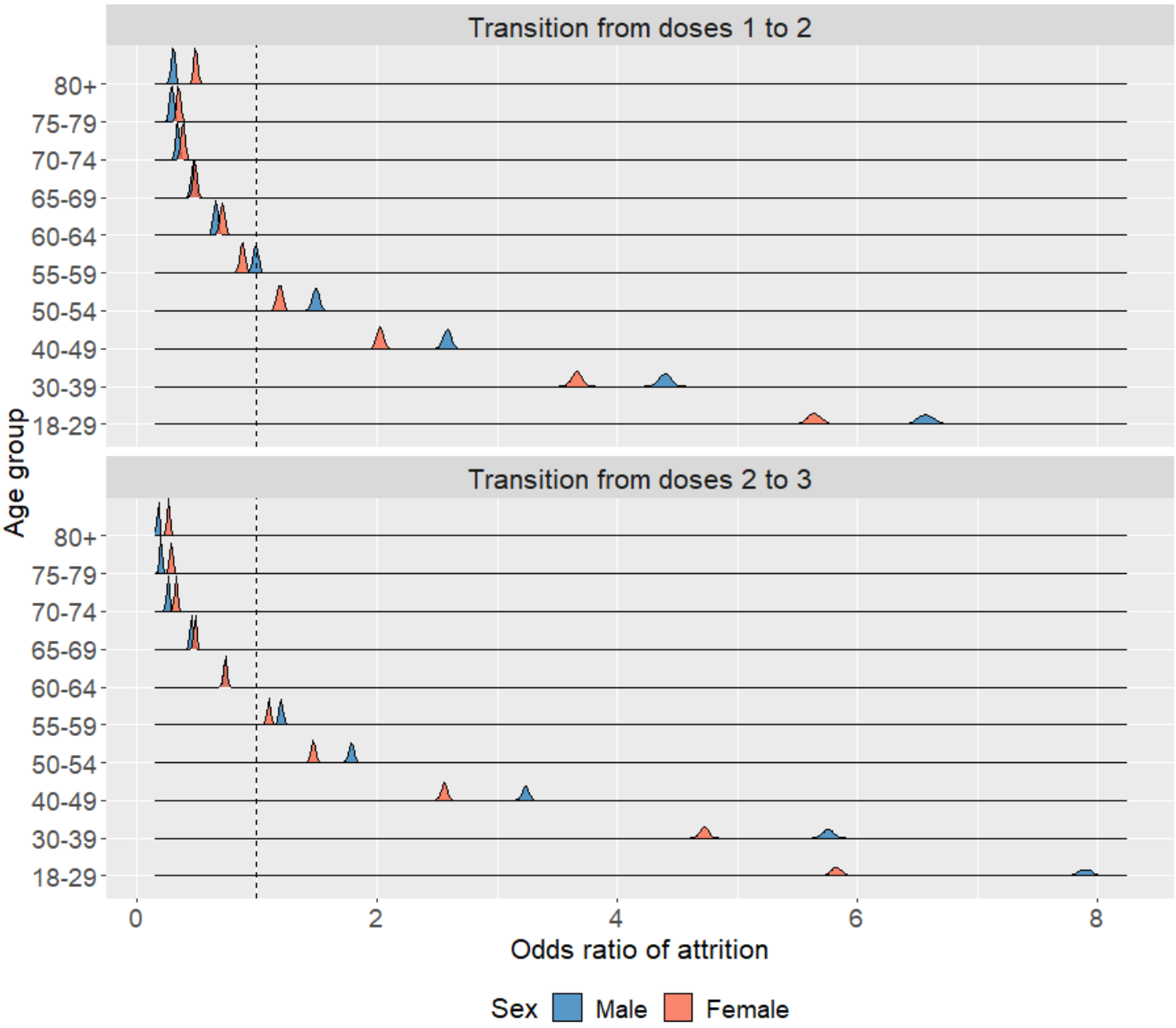


Figure 4.3: Posterior densities of the sex- and transition-specific odds ratios of attrition for each age group, compared to the average odds over all age groups.

Furthermore, the variation in the odds in favour of attrition is larger for males than females, with the posterior estimates being overall further away from the sex-specific average odds over all age groups (the dashed line). Therefore, the age group effects appear to differ by sex and should hence be estimated for the two sexes separately.

The next motivating question is whether or not the age group effects differ for the transitions from doses 1 to 2 and doses 2 to 3. Table 4.1 suggests that this is the case, as Model 1 with sex- and transition-specific age group effects has a substantially lower WAIC value than the model with only sex-specific age group

Table 4.2: Posterior mean estimates and 95% credible intervals of the sex- and transition-specific odds ratios of attrition by age group, compared to the average odds across all age groups.

Age group	Doses 1 to 2		Doses 2 to 3	
	Female	Male	Female	Male
80+	0.49 (0.46,0.52)	0.30 (0.28,0.33)	0.26 (0.25,0.27)	0.18 (0.17,0.19)
75-79	0.35 (0.32,0.38)	0.29 (0.27,0.31)	0.29 (0.27,0.30)	0.20 (0.19,0.21)
70-74	0.39 (0.36,0.41)	0.34 (0.32,0.36)	0.32 (0.31,0.34)	0.26 (0.25,0.27)
65-69	0.48 (0.46,0.51)	0.47 (0.44,0.50)	0.49 (0.47,0.50)	0.45 (0.44,0.47)
60-64	0.71 (0.68,0.75)	0.66 (0.63,0.69)	0.74 (0.72,0.76)	0.74 (0.72,0.76)
55-59	0.88 (0.84,0.91)	0.99 (0.95,1.03)	1.10 (1.08,1.13)	1.20 (1.17,1.23)
50-54	1.19 (1.15,1.23)	1.49 (1.44,1.54)	1.47 (1.44,1.50)	1.79 (1.75,1.82)
40-49	2.02 (1.97,2.08)	2.58 (2.52,2.65)	2.56 (2.52,2.61)	3.24 (3.18,3.30)
30-39	3.66 (3.58,3.75)	4.40 (4.30,4.50)	4.73 (4.65,4.81)	5.75 (5.66,5.85)
18-29	5.64 (5.52,5.76)	6.57 (6.43,6.71)	5.83 (5.73,5.92)	7.90 (7.76,8.03)

effects (with a WAIC value of 17,930.90 compared to 18,937.93). In Figure 4.3, the overall trends in the sex-specific odds ratios across age groups seem relatively similar in the two transitions. However, it should be noted that for each sex, for age groups 59 years and younger the odds ratios are consistently higher in the second transition than in the first. In contrast, for each sex, for age groups 70 years and older, the odds ratios are consistently smaller in the second transition than in the first. The numerical summaries in Table 4.2 confirm these observations. Hence, for both sexes, the variation in the odds of attrition between age groups is larger in the second transition than in the first. These findings confirm the previously supposed presence of an interaction between age group and transition in the attrition rates and that age group effects should be estimated for the two transitions separately.

4.4.2 Are there any spatial patterns in attrition rates by council area, and does this vary by sex or transition?

Here, the main question is if there are any spatial patterns (at a council area level) in the attrition rates. Table 4.1 shows that including the random council area effects in the model reduces the WAIC value substantially, but only after including the random age group and fixed transition effects. Thus, while the attrition rates seem to differ by council area, the impact of the spatial effects on the attrition rates appears to be more subtle than that of the age group and transition effects. In contrast to the age group effects, based on the WAIC and DIC values, it is unclear whether the

council area effects should be estimated separately for the two sexes. Model 1, with transition- but not sex-specific council area effects, has the lowest WAIC value of all models considered. However, the lowest DIC value is obtained for Model 2, which has sex- and transition-specific council area effects. Since there is not a clear best model, this section presents the results for both Model 1 and Model 2. For Model 1, the estimated posterior mean of ρ_ϕ is 0.414 with a 95% credible interval of (0.232, 0.700). For Model 2, the estimated posterior mean of ρ_ϕ is 0.337 with a 95% credible interval of (0.144, 0.579). Since the estimates of the mixing parameter are not close to zero for either model, a moderate amount of the variability in the council area effects is accounted for by the correlated component.

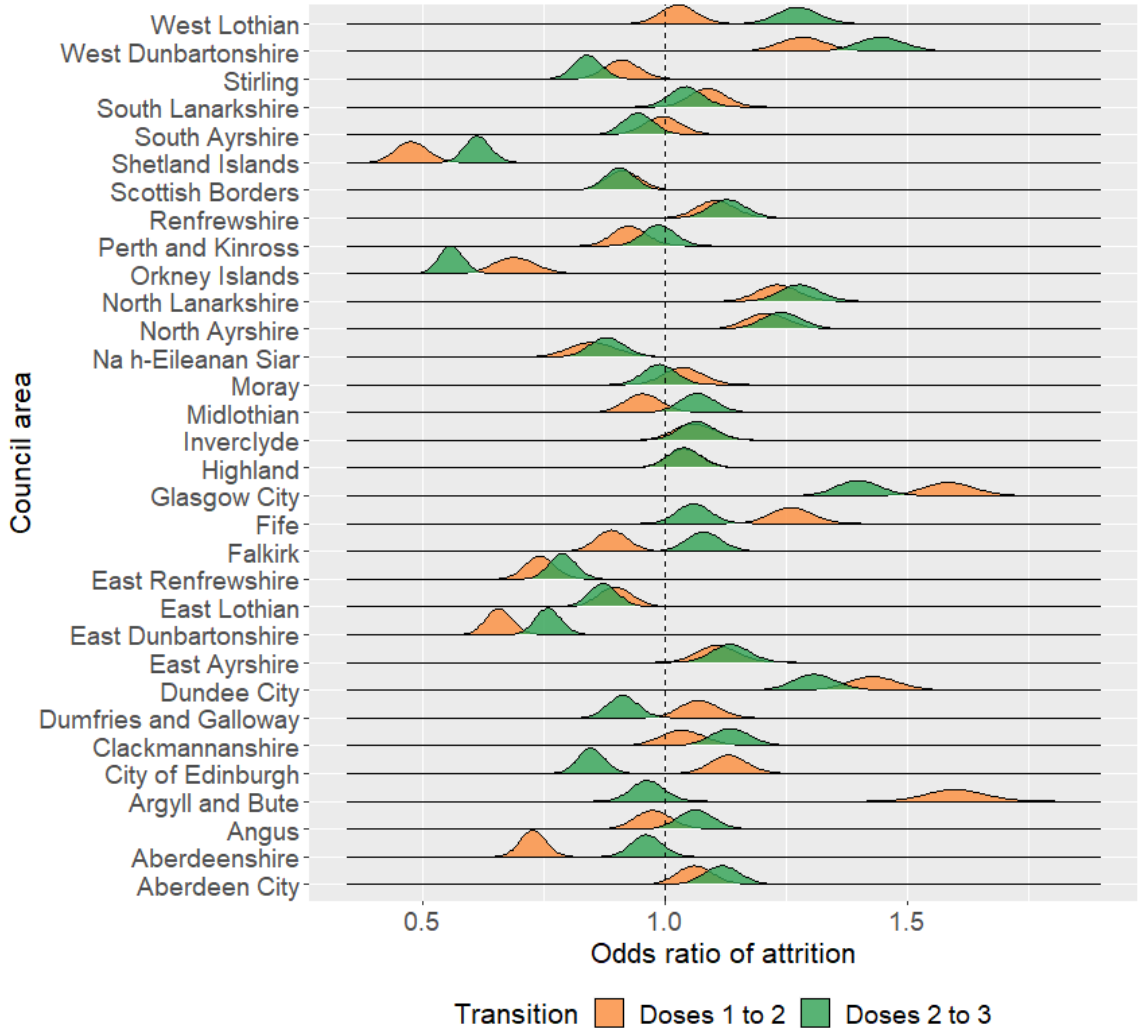


Figure 4.4: Posterior densities of the transition-specific odds ratios of attrition for each council area, compared to the Scottish average over both sexes.

Figure 4.4 shows posterior densities of the odds ratios of attrition $\left\{\exp(-\phi_k^{(d)})\right\}$ for council area k, compared to the corresponding transition-specific Scottish average

odds over all council areas, obtained from Model 1. The dashed line represents an odds ratio of one, i.e. when the transition-specific odds of attrition in a council area equal the Scottish average over both sexes. Table B.5 in the appendix provides the corresponding posterior means and 95% credible intervals. Over the two transitions, Glasgow City, West Dunbartonshire, and Dundee City are the council areas with the highest odds of attrition in Scotland. The council areas with the lowest odds are the Shetland Islands, the Orkney Islands, East Dunbartonshire, and East Renfrewshire. There appear to be substantial differences in the odds of attrition by council area, which confirms the previous suggestion that the model should contain spatial random effects. Moreover, for some of the council areas, the transition-specific densities of the odds ratios overlap barely (if at all), implying that the odds in these council areas changed differently compared to other council areas in Scotland from one transition to the next. The most prominent differences can be observed for Aberdeenshire, Argyll and Bute, the City of Edinburgh, Dumfries and Galloway, East Dunbartonshire, Falkirk, Fife, Glasgow City, Midlothian, the Orkney Islands, the Shetland Islands, West Dunbartonshire, and West Lothian. For example, in the City of Edinburgh, the odds in favour of attrition are higher than the Scottish average in the first transition (from doses 1 to 2), but smaller in the second transition (from doses 2 to 3). The odds in favour of attrition in Glasgow City are higher than the Scottish average in both transitions, but the odds ratio is smaller in the second transition than in the first. That is, compared to the transition-specific average odds over all council areas, the higher odds in Glasgow City appear to be less extreme in the second transition than in the first. These differences in the odds ratios of attrition between the two transitions confirm the previously supposed presence of an interaction effect between the variables council area and transition on the attrition rates and that the council area effects should be estimated for the two transitions separately. Lastly, it should be noted that the posterior densities of the odds ratios of attrition tend to be wider in the transition from doses 1 to 2 than in the transition from doses 2 to 3. The greater uncertainty in the odds ratios could be due to the odds in favour of attrition generally being much smaller in the first transition.

Figure 4.5 shows posterior densities of the sex- and transition-specific odds ratios

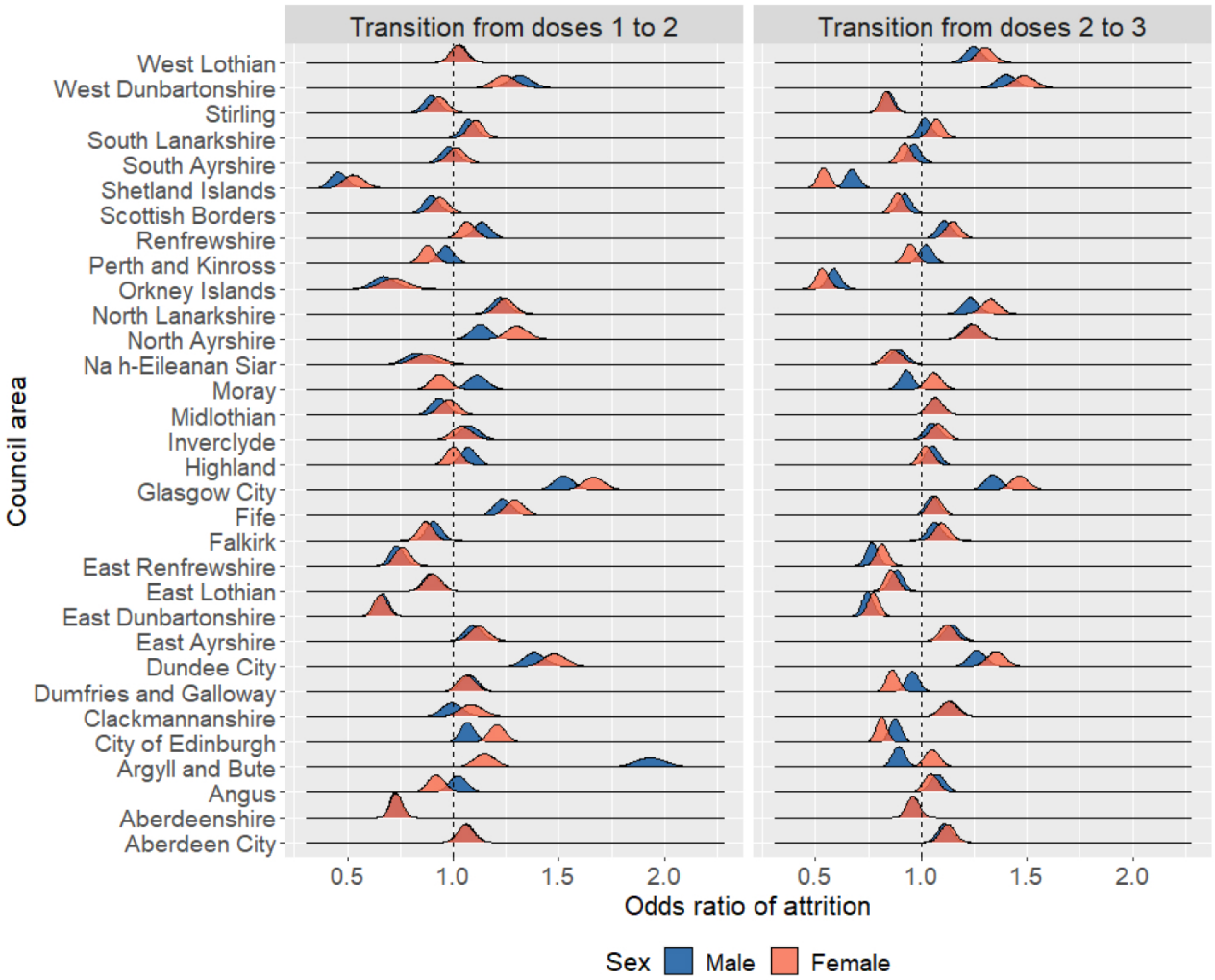


Figure 4.5: Posterior densities of the sex- and transition-specific odds ratios of attrition for each council area, compared to the Scottish average.

of attrition $\left\{\exp(-\phi_k^{(sd)})\right\}$ for council area k, compared to the Scottish average, obtained from Model 2. The corresponding posterior means and 95% credible intervals are presented in Table B.6 in the appendix. For each council area, taking the averages of the estimated posterior means of the sex- and transition-specific odds ratios over both sexes results in very similar estimates to the corresponding transition-specific posterior mean estimates obtained from Model 1, with a mean absolute difference of 0.004. The five largest absolute differences are observed for Argyll and Bute in the transition from doses 1 to 2 (0.058), the Shetland Islands in the transition from doses 1 to 2 (0.017), Moray in the transition from doses 1 to 2 (0.012), Argyll and Bute in the transition from doses 2 to 3 (0.009), and Glasgow City in the transition from doses 1 to 2 (0.007). Overall, the mean absolute differences between the average estimated posterior means of the sex- and transition-specific odds ratios over both sexes from Model 2 and the transition-specific odds ratios from Model 1 are

greater for the transition from doses 1 to 2 (0.006) than the transition from doses 2 to 3 (0.002). Much of this difference is explained by the large difference between the sex-specific odds ratios of Argyll and Bute in the transition from doses 1 to 2. However, computing the median absolute differences (which won't be affected by the unusual observation of Argyll and Bute) between the average estimated posterior means of the sex- and transition-specific odds ratios over both sexes from Model 2 and the transition-specific odds ratios from Model 1 shows that there are generally larger differences between the random council area effect estimates from Models 1 and 2 for the transition from doses 1 to 2 (0.003) than the transition from doses 2 to 3 (0.001).

Overall, the differences in the odds ratios between the two sexes in Figure 4.5 appear to be less pronounced than the differences between transitions in Figure 4.4. Nonetheless, for some council areas in Figure 4.5, there are noticeable differences in the odds ratios of attrition between the two sexes. For example, in Glasgow City and Dundee City, the odds ratios of attrition tend to be greater for females than males in both transitions. Thus, the higher odds in Glasgow City and Dundee City appear more extreme for females than males, compared to the corresponding sex-specific average odds in Scotland. Furthermore, while the odds ratios are greater than one for both sexes in both transitions, they appear less extreme in the second transition than in the first for both sexes.

In the City of Edinburgh, in the first transition, for both sexes the odds in favour of attrition are higher than the transition- and sex-specific Scottish average odds, with females having a larger odds ratio than males; i.e., in the first transition, the higher odds of attrition for females deviate more from the sex-specific national average than those of males. In contrast, in the second transition, for both sexes in the City of Edinburgh the odds in favour of attrition are lower than the sex-specific Scottish average odds, where the odds ratio is likely to be smaller for females than males; i.e., the lower odds of attrition for females appear to deviate again more from the sex-specific national average than those of males. These findings reflect a possible interaction between sex and transition in the council area effects on attrition. Another example for such an interaction is Moray, where the posterior density for the female

odds ratio is mostly below one in the first transition and likely above one in the second, while the male odds ratio is mostly above one in the first transition and likely below one in the second. The most extreme case is that of Argyll and Bute in the transition from doses 1 to 2, where for each of the two sexes the odds in favour of attrition are higher than the corresponding transition- and sex-specific Scottish average odds, but the odds ratios are much larger for males than females, with posterior mean estimates of 1.93 and 1.14, respectively. In the second transition, the odds in favour of attrition for males in Argyll and Bute are 0.89 and hence, the odds are smaller than the transition- and sex-specific Scottish average odds, while those of females are 1.05 and hence higher than their corresponding national average. The extremely large odds ratio for males in Argyll and Bute in the first transition might explain why the odds ratio is below one in the second transition; i.e., with an unusually large proportion of males refusing the second dose, the relatively small proportion that did accept the second dose might have been generally more inclined to continue the vaccination programme.

4.5 Discussion

While vaccine hesitancy (the delay in acceptance or refusal of vaccination despite the availability of vaccination services) for COVID-19 has previously been studied extensively, the acceptance of later doses appears to have been considered only for the booster vaccine, and it seems that vaccine fatigue (the delay in acceptance or refusal of subsequent vaccine doses) is otherwise unexplored. Hence, the motivation for this study was to analyse possible trends and patterns in the attrition rates between subsequent doses of the COVID-19 vaccine, aiming to identify differences in attrition rates by sex, age, area, and transition between doses. The study was based in Scotland, and the cumulative number of first, second, and third doses of the COVID-19 vaccine up to 19th April 2022 was recorded at a council area level by sex and age group. A binomial logistic model was fitted to these attrition rate data in a Bayesian setting using the `R-INLA` package. The sex- and transition-specific random council area and age group effects were modelled with flexible BYM2 models that allow for

structural autocorrelation, and possible interactions were considered between age and sex, age and transition, area and sex, and area and transition.

The main finding is that the fitted models show a strong association between age and attrition rates, where the odds in favour of attrition decrease smoothly with increasing age (i.e., older age groups are more likely to continue receiving vaccine doses). This association between older age and lower attrition rates is consistent with previous findings of older people being less hesitant to receive a first dose (Soares et al., 2021, Mewhirter et al., 2022) and having a higher willingness to receive a booster vaccine (Rzymski et al., 2021, Sønderskov et al., 2021). Since older individuals were prioritised and their higher risk of severe disease progression was emphasised prior to the initial vaccine rollout in Scotland (Scottish Government, 2021a), it seems plausible that the more severe potential consequences of not having the vaccine results in lower attrition rates for the older age groups.

In this study, the age group effects are allowed to differ by sex and transition. The resulting estimates suggest that the age group effects differ for the two sexes, with the overall variation in the odds of attrition between age groups being greater for males than females. Compared to the sex-specific average odds over all age groups, the odds ratios of attrition are higher for males than females in age groups 59 years and younger. This trend reverses for age groups 70 and older, where the odds ratios are below one for both sexes but higher for females than males. Thus, the decreasing effect of older age on the odds of attrition appears to be smaller for females than males, relative to the corresponding sex-specific average odds over all age groups. This finding aligns with the observed attrition rates tending to be higher for females than males in age groups 70 years and older (see Figure 4.1). The age group trends are similar for the two transitions, although the variation in the odds of attrition between age groups tends to be overall greater in the second transition than in the first. Comparing these findings with other studies is impossible as this study appears to be the first to consider such interactions between age and sex and between age and transition. The identified interactions between age, sex, and transition suggest that other studies on COVID-19 vaccine uptake should also consider the age group effects to be sex- and transition-specific, if applicable.

Additionally, the best fitting models in this study allow estimation of the overall differences in attrition rates between the two sexes. These findings show that for a fixed transition, averaged over all age groups and council areas in Scotland, males tend to have approximately 1.1 times higher odds of attrition than females. For vaccine hesitancy, the differences by sex have not been consistent across studies, with some showing that the acceptance of a first dose is lower for females than males (Troiano and Nardi, 2021), and others indicating that males have a lower acceptance (Lazarus et al., 2021). The same applies to the booster vaccine willingness, where some studies found that females are associated with higher willingness than males to receive the booster dose (Rzymski et al., 2021), while others showed a lower willingness in females (Gaffney et al., 2022). Therefore, the finding of overall greater vaccine fatigue in males in Scotland might not necessarily be replicated in other countries.

For a fixed sex, the average odds of attrition over all age groups and council areas in Scotland are 4.1 times higher in the transition from doses 2 to 3 than from doses 1 to 2, which suggests increased vaccine fatigue over time. Note that the two primary doses of the COVID-19 vaccine were recommended from the start of the vaccine rollout in December of 2020 in Scotland (Scottish Government, 2020a), while the Joint Committee on Vaccination and Immunisation (JCVI) first advised a potential booster vaccine on 30th June 2021 (Department of Health & Social Care, 2021a). Thus, individuals might have been more likely to refuse the booster vaccine as the additional third doses were advised later. Furthermore, in Scotland, regulations came into force on 1st October 2021 (Swinney, 2021), requiring a person to be fully vaccinated (having received the first and second doses), unless exempt, to attend most events. This mandate might have motivated individuals to receive the first and second doses of the vaccine. The mandate did not apply to the third dose, which might have contributed to the larger attrition rates in the transition from doses 2 to 3. However, vaccine uptake was already high when the mandate came into force on 1st October 2021, with 4.06 million adults in Scotland (91.44% of the approximate total adult population of 4.44 million people) having received the first dose and 3.74 million (84.23%) having received the first and second doses of the COVID-19 vaccine (UK Health Security Agency, 2023b). A slight increase in the daily administered first

doses of vaccine was observed around the advent of the regulations, whereas the daily number of second doses was at a constant level around that time. Hence, it seems unlikely that the government mandate is a leading cause for the much higher odds of attrition for the second transition (from doses 2 to 3) compared to the first (from doses 1 to 2). Another possible explanation for the increased odds of attrition in the second transition is that some individuals might have been infected with COVID-19 after they received the second dose and no longer believed they had a need to receive a booster vaccine. Rzymski et al. (2021) and Dziedzic et al. (2022) add weight to this hypothesis, having identified a reduced willingness to receive the booster vaccine in the Polish population previously infected with the virus. It should be noted that the NHS advised to wait four weeks from the date of a positive test or first symptoms to get the vaccine (Scottish Government, 2022), which might have lead to those infected at the time of vaccine invitation to not book an appointment after the four weeks had passed.

The final main finding is that over the two transitions and both sexes, Glasgow City, West Dunbartonshire, and Dundee City are the council areas with the highest odds of attrition in Scotland. The council areas with the lowest odds are the Shetland Islands, Orkney Islands, East Dunbartonshire, and East Renfrewshire. It should be noted that Glasgow City, West Dunbartonshire, and Dundee City are amongst the five most deprived council areas in Scotland, while the Shetland Islands, Orkney Islands, East Dunbartonshire, and East Renfrewshire are amongst the ten least deprived council areas (Scottish Government, 2020b). Hence, it seems possible that deprivation, which was previously identified to be associated with greater vaccine hesitancy (Kumari et al., 2021, Lee and Huang, 2022) and greater unwillingness to receive the booster vaccine (Paul and Fancourt, 2022) is generally a driving factor for the higher odds in favour of attrition between doses.

Some data limitations are apparent in this study. While the deceased and those who no longer live in Scotland are not included in the vaccination data, it is possible that some people moved home to a different council area, changed age groups, or became ineligible between subsequent doses of the vaccine, which will have affected the results to a small extent. While this study analysed sex, transition, age group, and

council area effects on attrition rates, many other variables previously linked to vaccine uptake were not considered because the data were aggregated to a relatively high council area level, meaning that covariate effects could be impacted by ecological bias (Wakefield and Salway, 2001). Furthermore, there are only 32 council areas, so there is a risk that any regression relationship will be badly estimated. To examine these concerns, I refitted the final model with the additional covariates income deprivation and population density and provide the results in Section B.10 of the appendix. The resulting fitted attrition rates are almost identical to the ones of Model 2 (from Table 4.1), so the results are not sensitive to the omission of these additional covariates. Another variable that could impact attrition rates is the type of vaccine received in an earlier dose, as Dziedzic et al. (2022) suggested that those who had previously been vaccinated with Pfizer/BioNTech or Moderna had higher booster uptake than those vaccinated with AstraZeneca. However, the type of vaccine administered was not specified in the data and hence it could not be considered in this study.

The BYM2 model assigned to the council area effects in this study allows for spatial autocorrelation in the attrition rates across council areas. However, assigning a binary adjacency neighbourhood structure to the BYM2 model forces the correlated effects to smooth the attrition rates strictly across council areas that share a border, which might not be appropriate in all cases. For example, Glasgow City and East Dunbartonshire share a border but their attrition rates differ greatly, with Glasgow City having an average observed attrition rate of 11.47% across the two transitions, which is the highest in Scotland, compared to 5.27% in East Dunbartonshire, which is the lowest in the Scottish mainland. Thus, rather than assuming smoothness in the attrition across all neighbouring council areas, the model could be improved by locally smoothing the attrition rates (e.g., see Lee and Mitchell, 2013). While the risk surface of this study is limited to 32 council areas, having 40 observations for each council area might be sufficient to identify localised spatial structures in the data.

The model's random age group and council area effects are allowed to differ by sex and transition. However, they are assumed to have the same variance and mixing parameters to allow for more observations in estimating them. Since the variation in the modelled attrition rates was greater for males than females and greater in

the second transition than the first, I considered allowing sex- and transition-specific hyperparameters. However, this resulted in a worse model fit, with WAIC and DIC values exceeding those of the original Model 2 by more than 2,000 each, likely because there are such few data (10 age groups and 32 council areas for each combination of sex and transition). Allowing the hyperparameters to differ for these sub-groups might be a viable option in studies with a larger number of observations.

A profitable avenue for future work, if the data were available, is to examine attrition rates for different types of vaccines that require multiple doses (e.g., two doses are recommended for MMR vaccines, two doses for varicella (chickenpox) vaccine, and annual doses for influenza vaccine). Another possibility is to extend the study to a multi-country comparison to see if the trends and patterns in the attrition rates in the COVID-19 vaccination programme in Scotland are also found in other countries or to see how they differ. Lastly, if the study was conducted on smaller areal units, analysing the relationship between attrition rates and other covariates such as income deprivation rate, population density, and education levels would provide interesting additional insights.

The presented study featured the unusual observation of a very high attrition rate for males in Argyle and Bute in the transition from doses 1 to 2 (see Figure 4.2). When the data are overall strongly positively spatially autocorrelated, such unusual observations can be difficult to identify from the smoothed estimates of a spatial smoothing model and they might affect the inference on the smooth risk or prevalence surface. Hence, motivated by the unusual observation identified in this study, Chapter 5 provides a study of how such unusual observations might be identified in big data (with thousands of areal units) and how a smoothing model could be modified to remove the unusual observations' impact on the estimated disease prevalence surface.

Chapter 5

When Tobler's First Law of Geography doesn't hold: Identifying spatially outlying observations to remove their impact on estimated disease prevalence surfaces

5.1 Introduction

Publicly available disease incidence (or prevalence) data are most often spatially aggregated into small areas, with population sizes sufficiently large to protect confidentiality but small enough to represent near-homogeneous communities. However, most diseases are relatively rare when viewed at a small area population level, and hence, these data are subject to substantial random variation. Rather than working directly with the noisy observed data, disease mapping models infer the latent risk (or prevalence) surface, which is commonly assumed to change smoothly across neighbouring areas, and aim to capture the overall pattern in disease risk (e.g., see

MacNab, 2022 for an extensive review). As a result, regions with unusual risks can be identified from these maps, such as high-risk regions that may require public health intervention. However, unusual risks in individual areas are usually smoothed over in the modelling process and hence can be difficult to identify (e.g., see Richardson et al., 2004).

The assumption of spatial smoothness is motivated by Tobler's *First Law of Geography*, that "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). In epidemiology, underlying features that could result in a spatially smooth risk surface include latent risk factors (e.g., air pollution) that change smoothly in space, grouping effects due to neighbourhoods forming based on population similarities (e.g., ethnicity, income), and the behaviour of people living close and interacting with each other becoming more similar over time. Some disease mapping models strictly enforce spatial smoothness in the inferred risks, while others estimate the degree of spatial autocorrelation via a spatial smoothing parameter. When the spatial smoothing parameter is estimated as a single value, it implicitly assumes that the strength of spatial autocorrelation in disease risk is similar across the entire study region, resulting in a globally smooth risk surface.

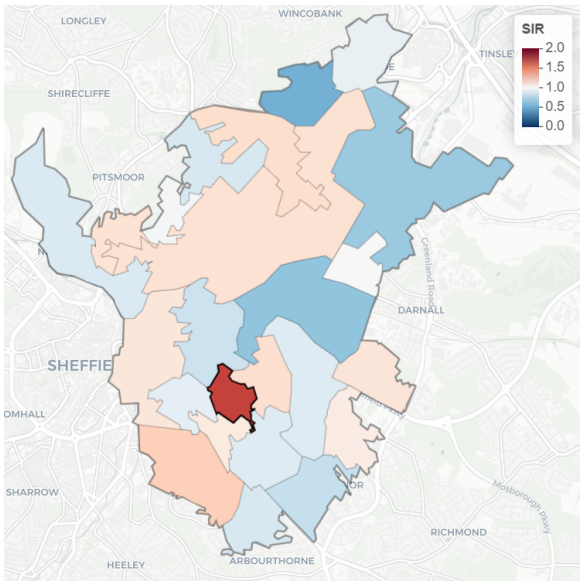
However, structural deviations from the global smoothness assumption for disease risk often exist, and can be due to physical boundaries, such as hills, rivers, and highways that prevent communities from mixing, as well as social frontiers (e.g., see Dean et al., 2019) where social/ethnic characteristics differ substantially between neighbouring communities. When the global spatial smoothness assumption is not appropriate, adaptive smoothing models can be applied (e.g., see Lu and Carlin, 2005, Lee and Mitchell, 2013, Rushworth et al., 2017, Gao and Bradley, 2019, Corpas-Burgos and Martinez-Beneito, 2020) to allow for local discontinuities in the estimated risk surface between neighbouring areas. Although the adaptive smoothing models provide better disease risk estimates than global smoothing models, their estimated risk surfaces will still be locally smooth and may be adversely affected by individual areas with unusual risks that break with the otherwise locally smooth risk patterns. Therefore, individual areas with unusual outcomes should be identified before fitting a smoothing model so that one can check the impact of these observations on the

statistical inference. With that motivation in mind, this study proposes a spatial outlier detection method for areal data to identify individual areas that violate the local smoothness assumption, followed by a modified smoothing model that accounts for these potential outliers. To distinguish these individual outlying areas from hotpots or contiguous clusters where the observations in more than one area appear unusual, the remainder of this chapter will refer to the individual outliers as *singleton outliers*.

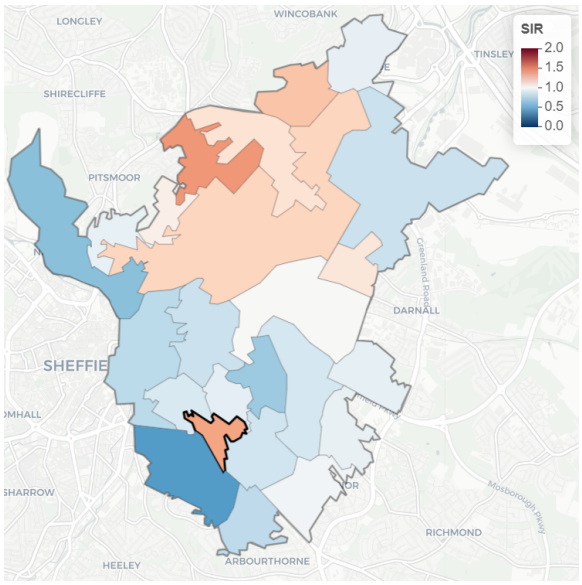
An outlying observation, or *outlier*, has historically been defined as an observation that "appears to deviate markedly from other members of the sample in which it occurs" (Grubbs, 1969). To distinguish the considered outliers from the rest of the data, in this study, those observations that are not singleton outliers are referred to as *inliers*. In experimental studies, two possible causes could lead to outlying observations. One is that these observations result from the random variability in the data-generating process, where, by chance, some outcomes are extreme. Alternatively, outliers could result from data collection or processing errors. In applications that use observational data, many other possible causes for outliers exist. For example, in public health analyses, environmental factors or the presence of related individuals with a rare genetic disease could explain unusually high disease risks in some small areas.

Chandola et al. (2009) distinguish between three types of outliers: global outliers (type I), contextual outliers (type II), and collective outliers (type III). Here, the focus is on small-area disease mapping applications, so Figure 5.1 presents examples of these three types of outliers in standardised incidence ratio (SIR) data for an example set of small areas (the data used to create these maps are randomly generated and presented at the English lower super output area level around Sheffield, UK). Global outliers are individual observations inconsistent with the rest of the dataset and, therefore, fall under the previously mentioned definition by Grubbs (1969). For the SIR example, a global outlier is a single areal unit with an SIR that differs markedly from the rest of the data, such as the one highlighted with a thicker border in Figure 5.1a. A contextual outlier is an individual observation that is an outlier within its local spatial context. That is, a contextual outlier is considered part of the inlying data when viewing the dataset as a whole but does not conform to the typical local behaviour.

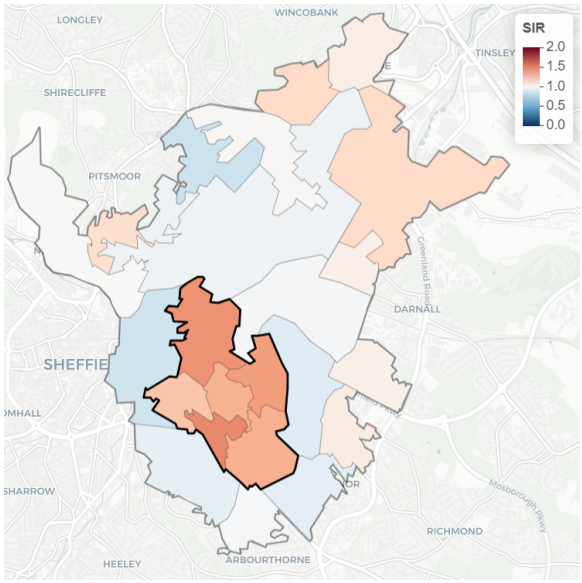
For example, the highlighted area in Figure 5.1b is a contextual outlier with a higher SIR surrounded by areas with lower SIRs. Collective outliers are a subset of data instances whose collective occurrence is unusual. In the framework of this study, collective outliers are clusters of spatially contiguous areas (any area in the cluster shares a border with at least one other area in the cluster). The highlighted cluster in Figure 5.1c forms a set of collective outliers as its SIR values are higher than those in the areas surrounding the cluster.



(a) Global outlier (type 1)



(b) Contextual outlier (type 2)



(c) Collective outlier (type 3)

Figure 5.1: Example maps showing the three types of spatial outliers.

Several general methods for identifying spatial clusters (collective outliers) have been proposed (e.g., see Murray et al., 2014, Campello et al., 2020), while applications in epidemiology can be found, for example, in Huang et al. (2008) or Li et al. (2019).

In contrast, relatively few studies developed spatial outlier detection techniques for global or contextual outliers (i.e., singleton outliers). Some studies have proposed methods for point-referenced or geostatistical data (e.g., Shekhar, 2003, Adam et al., 2004, Birant and Kut, 2006, Sedaghat et al., 2013, Duggimpudi et al., 2019), while Cheng and Li (2006) analyse images over a regular grid. The current study focuses on identifying potential singleton outliers (global or contextual) on a map of irregularly shaped areal units. In this context, Anselin (1995) proposes local indicators of spatial association (LISA) for hotspot and outlier detection, using a local version of Moran's I statistic (Moran, 1950). For a review of Moran's I statistic and Anselin's local Moran's I statistic, revisit Section 2.5.1. A negative local Moran's I statistic points to an area with a value different from its neighbouring values, i.e., an area that appears to be a contextual or global outlier. The local Moran's I statistic is an established method in spatial statistics textbooks (e.g., see Moraga, 2023) for identifying possible outliers and, hence, identifying singleton outliers using local Moran's I statistic is used as a comparison to this study's novel methodology.

As a general principle, local spatial outlier detection methods define neighbourhoods based on the data's spatial attributes and identify outliers whose non-spatial attributes deviate considerably from their neighbouring values. Schubert et al. (2014b) provide a general review of local outlier detection, including methods that have been proposed for areal data, such as the works of Lu et al. (2003), Kou et al. (2006), and Chawla and Sun (2006), who examine their methods on US county-level data. Chen et al. (2008) propose an improved version of the median algorithm by Lu et al. (2003), but similar to the previous studies, they only evaluate their method using illustrative examples. In contrast, Singh and Lalitha (2018) propose an improved version of the mean and median algorithms by Lu et al. (2003) and demonstrate the improved performance using simulated datasets. However, the resulting receiver operator characteristic (ROC) curves and area under the curves (AUCs) indicate that the method does not perform substantially better than a random classifier. Schubert et al. (2014a) propose a localised method that uses kernel densities for outlier detection and they apply their method to spatial and spatio-temporal geostatistical data. Tang and He (2017) propose an alternative kernel density-based outlier score,

which they evaluate on synthetic two-dimensional point data and multidimensional non-spatial data. For the latter, they show that their method outperforms that of Schubert et al. (2014a). Neither study investigates their method's sensitivity to the choice of bandwidth parameter value, which has a large impact on the kernel densities and should be analysed further.

The study presented here proposes a novel two-step disease mapping approach that accounts for possible singleton spatial outliers. Potential singleton spatial outliers are identified in the first step, and the proposed method extends the non-parametric approach of Tang and He (2017) to spatial locations with non-spatial attributes. The method is evaluated in simulation study and compared to a singleton spatial outlier detection using local Moran's I, which is commonly used to identify possible outliers in areal data. For the second step of the method, a modified disease mapping model is proposed that accounts for the potential singleton spatial outliers identified in the first step. The efficacy of the proposed model is again evaluated in simulation study and compared to a conventional smoothing model that smoothes over all areas.

The development of the method is motivated by a novel study of asthma prevalence at the lower super output area (LSOA) level in England. In this study, potential singleton spatial outliers are identified for the year 2017 (the most recent year in the data), allowing public health experts to identify the drivers of these outlying observations. The observed number of asthma cases are fitted with a binomial generalised linear model using the proposed modified smoothing model that accounts for the potential outliers, and the resulting estimates are compared to those of a commonly used smoothing model that does not consider potential outliers. The study's aim is to answer the following three main questions:

1. Which areas appear to be potential singleton outliers?

2. Do the identified outliers appear to be contextual or global?

3. How do the potential outliers affect the modelling?

The remainder of this chapter is organised as follows. Section 5.2 presents the asthma prevalence data that motivate this study. Section 5.3 presents the novel

methodology: a relative density-based singleton outlier detection technique for small area data and a modified smoothing model that accounts for potential singleton spatial outliers to remove their impact on the estimated disease prevalence surfaces. The method's efficacy is assessed in simulation study in Section 5.4, and Section 5.5 applies the method to the motivating study presented in Section 5.2. Lastly, Section 5.6 provides the conclusions from this study and discusses limitations and avenues for future work.

5.2 Motivating study: Asthma prevalence at a small-area level in England

This study applies the proposed singleton spatial outlier detection method and modified smoothing model to small-area asthma prevalence data to identify individual areas whose asthma prevalence appears to differ from the typical local behaviour so that these areas can be identified and their underlying prevalence be estimated independently from the rest of the data to ensure that these unusual observations do not corrupt the estimation of the otherwise smooth prevalence surface. A singleton spatial outlier could have an asthma prevalence that differs from the rest of the dataset as a whole (i.e., be a global outlier), have a higher asthma prevalence and be located in a region that tends to have a lower prevalence, or have a lower asthma prevalence and be located in a region that tends to have a higher prevalence (i.e., be a contextual outlier). The individual outlying areas are of interest to public health experts since areas with an unusually high asthma prevalence could benefit from public health interventions, and areas with an unusually low prevalence could be analysed to identify factors that might be associated with a lower asthma prevalence.

5.2.1 Study data

The Place-based Longitudinal Data Resource (https://pldr.org/dataset/e6nzv/quality-and-outcomes-framework-indicators-asthma-prevalence-qof403) provides publicly available annual asthma-prevalence data for all ages over the

years 2005 to 2017 at the lower super output area (LSOA) level in England, which partitions the country into 32,844 areas. For each LSOA and year, the observed asthma prevalence is measured as the estimated number of patients diagnosed with asthma divided by the total number of registered patients. These values are computed as weighted averages from general practitioner (GP) data, where any patient with an asthma diagnosis who has been prescribed asthma-related drugs in the preceding 12 months counts towards the asthma prevalence, although GP practices reporting asthma prevalence beyond 30% are excluded from the calculation. For this study, these values are rounded to the nearest integer, and only the most recent data from 2017 are analysed. For that year, the observed prevalences range from zero to 0.143.

5.2.2 Exploratory analysis

Let $\boldsymbol{y} = (y_1, \ldots, y_K)$ denote the observed number of patients with a diagnosis of asthma for 2017 in the LSOAs $(1, \ldots, K)$, where $K = 32,844$. Similarly, let $\boldsymbol{N} = (N_1, \ldots, N_K)$ denote the number of registered patients in all areas. The observed asthma prevalence in LSOA j is then computed as $p_j = y_j/N_j$, for $j = 1, \ldots, K$ and is a raw estimate of the true but unknown asthma prevalence. Figure 5.2 shows a histogram of the observed prevalence. The distribution is left-skewed, and the prevalence has a mean of 0.060, a median of 0.061, and a standard deviation of 0.010.

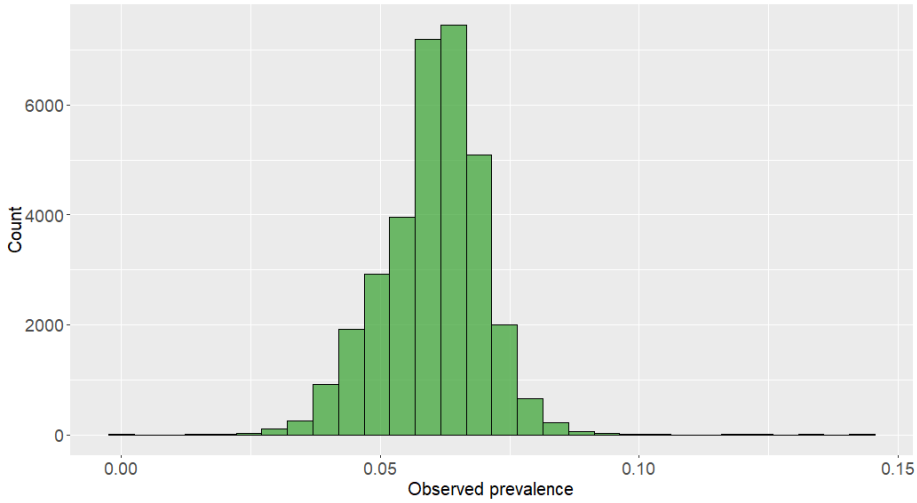


Figure 5.2: A histogram of the observed asthma prevalence in 2017.

As discussed in Section 4.1, public health data tend to change smoothly across the map, meaning that areas near each other tend to have similar health characteristics.

Moran's I statistic (recall Section 2.5.1.1) can be used to measure the spatial autocorrelation in the observed prevalences $\boldsymbol{p} = (p_1, \ldots, p_K)$, and a permutation test can be used to formally test for overall spatial autocorrelation in the data (e.g., see Lee, 2020). Here, the neighbourhoods are defined as the six-nearest neighbours measured by the Euclidean distance between the areas' centroids, and the neighbourhoods are made symmetric so that areas i and j are neighbours if area i is among the six nearest neighbours of area j, or vice versa. Using the six nearest is neighbours is motivated by each area in the study region sharing a border with an average of 5.87 areas. For these neighbourhoods, the Moran's I statistic for $\boldsymbol{p}$ is 0.89 and a permutation test with 10,000 permutations results in a p-value smaller than 0.0001. Therefore, the observed prevalences $\boldsymbol{p}$ are significantly positively spatially autocorrelated, meaning that neighbouring values tend to be more similar. Hence, the data are suitable for the proposed outlier detection method to identify whether there are individual areas with asthma prevalences that do not follow their typical local behaviour.

5.3 Method: Identifying spatially outlying observations to remove their impact on estimated disease prevalence surfaces

This section presents the two-step modelling approach proposed in this study. Section 5.3.1 presents the proposed method for identifying potential singleton (global or contextual) spatial outliers, and Section 5.3.2 presents a modified smoothing model to remove their impact on estimated disease prevalence surfaces.

5.3.1 Identifying spatial outliers

This section presents the proposed method for identifying potential singleton spatial outliers using a relative density-based outlier score (RDOS). The RDOS statistic is defined in Section 5.3.1.1, Section 5.3.1.2 provides a method for specifying the bandwidth parameter that is used for computing the scores, and Section 5.3.1.3 discusses how to choose the number of observations to be identified as outliers.

5.3.1.1 Relative density-based outlier score (RDOS)

For the framework of this study, assume the values $(y_1, \ldots, y_K)$ are observed in K areas labelled $(1, \ldots, K)$. Let $M_j = \{i : \text{ area } i \text{ is a neighbour of area } j\}$ denote the neighbourhood set of area j of size $|M_j|$. For example, one could apply the previously described k-nearest neighbour rule to define the neighbourhoods, or a border-sharing rule where two areas are defined to be neighbours if they share a border. The local kernel density of observation y_j is computed as

$$\tilde{f}_j(y_j) = \frac{1}{|M_j| + 1} \sum_{i \in (M_j \cup \{j\})} G_h(y_i - y_j), \tag{5.1}$$

where G_h is a kernel function and h is the smoothing bandwidth. In this study, G_h is defined to be a Gaussian kernel, computed as

$$G_h(y_i - y_j) = \frac{1}{(2\pi)^{1/2} h} \exp\left(-\frac{1}{2}\frac{(y_i - y_j)^2}{h^2}\right), \tag{5.2}$$

where h is a global smoothing bandwidth used in all neighbourhoods. The Gaussian kernel function $G_h(y_i - y_j)$ returns the density of a Gaussian distribution with mean y_j and variance h^2 evaluated at the value y_i. The density curve is symmetric and unimodal, with the mode located at the mean value y_j. Hence, the kernel function takes on a larger value when y_i is closer to y_j, and the bandwidth h controls how quickly the density decreases as one moves away from y_j. For a larger bandwidth h, the densities evaluated at two values y_i and $y_l = y_i + \delta$ for some small δ will be more similar to each other; i.e., the variability in the densities reduces as one increases the bandwidth h.

The relative density-based outlier score (RDOS) compares each observation's local kernel density to the average density of its neighbouring observations and is computed as

$$\text{RDOS}(y_j) = \left(\frac{1}{|M_j|} \sum_{i \in M_j} \tilde{f}_i(y_i)\right) \Big/ \left(\tilde{f}_j(y_j)\right), \tag{5.3}$$

where $\tilde{f}_j(y_j)$ denotes the local kernel density from Equation (5.1), for $j = 1, \ldots, K$.

The proposed outlier score from Equation (5.3) differs from that of Tang and He (2017) in the definition of the neighbourhood sets. Tang and He (2017) define the neighbourhood sets based on distances in the observed values, which can be either spatial point locations or non-spatial attributes. In this study, the neighbourhoods are formed based on the areas' spatial locations, and the kernel density of each observation is computed using the attributes observed in the neighbouring areas, which makes the RDOS from Equation (5.3) applicable to spatial data with non-spatial attributes.

Suppose RDOS(y_j) takes a value of one. In that case, the local kernel density of observation y_j is the same as the average local kernel density of its neighbouring observations $\{y_i : i \in M_j\}$. Similarly, when RDOS(y_j) is smaller than one, the local kernel density of observation y_j is larger than the average local kernel density of its neighbouring observations. In these cases, the observation y_j appears to follow the typical local behaviour. In contrast, when RDOS(y_j) is greater than one, the local kernel density of the observation y_j is smaller than the average local kernel density of its neighbouring observations. In that case, the observation y_j is overall more dissimilar from its neighbouring observations $\{y_i : i \in M_j\}$ than these observations are from their neighbouring observations, on average. Thus, the larger the value of RDOS(y_j), the more likely it is that area j is a singleton outlier.

When the RDOS values are ranked, the highest-ranked objects are most likely to be singleton spatial outliers. Note that the RDOS values do not indicate whether an area is a contextual or a global outlier. However, while global outliers appear unusual when viewing the whole dataset, contextual outliers appear unusual only relative to their neighbouring observations. Hence, the attribute values of the identified outliers can be compared to the whole dataset to check which ones appear global.

5.3.1.2 Bandwidth parameter specification

In this study, the bandwidth parameter is computed as a function of the local variability in the data. In this regard, the approach is similar to that of Schubert et al. (2014a). However, Schubert et al. (2014a) propose using neighbourhood-specific local kernel bandwidths. For the notation used here, the local bandwidth would be the average difference between an area's observation and its neighbouring observations

or a somewhat arbitrary upper bound ϵ to make the method more robust to outliers. To eliminate the need to specify an upper bound, here, a global bandwidth is computed as a scalar multiple of the median of the local measures of variability. Furthermore, the median absolute difference is used instead of the average absolute difference to measure local variability, reducing the impact of outliers. Specifically, the global bandwidth in this study is computed as

$$h_c = c \times \underset{j=1,\ldots,K}{\text{median}} \left(\underset{i \in M_j}{\text{median}}(|y_i - y_j|) \right), \text{ for } c > 0. \tag{5.4}$$

As previously described, the variability in the local kernel densities decreases as one increases the bandwidth h. As a results the variability in the RDOS values is also smaller for larger bandwidths. Therefore, increasing the bandwidth reduces the risk that an area has an unusually large RDOS value by chance and for larger bandwidths h and $\tilde{h} = h+\delta$ for some small δ, the RDOS rankings should be mostly the same; only the differences between the RDOS values are expected to decrease as one increases the bandwidth further. Hence, an initial bandwidth is computed as in Equation (5.4) for $c = 1$ and is increased incrementally until the rankings under subsequent bandwidths are mostly the same. If the rankings are the same between $c = 1$ and the next larger value, one could consider reducing the value of c below one. The smallest bandwidth for which the rankings stabilise is used in the final analysis, resulting in the largest differences between the RDOS values out of the bandwidths for which the rankings are stable. The rankings under different bandwidths are compared using Kendall's rank coefficient.

Specifically, for the current value c, the RDOS values are computed and ranked in decreasing order. Let $\boldsymbol{r} = (1, \ldots, K)$ denote the ranks. The scalar c is increased to $\tilde{c}$, and the RDOS values are again computed and ranked in decreasing order. Let $\tilde{\boldsymbol{r}}$ denote the new ranking for $\tilde{c}$, as a potential permutation of the previously computed ranks $\boldsymbol{r}$. Kendall's rank coefficient is then computed as

$$\tau = \frac{2}{K(K-1)} \sum_{i<j} \text{sgn}(r_i - r_j)\text{sgn}(\tilde{r}_i - \tilde{r}_j), \tag{5.5}$$

where the function $\text{sgn}(\cdot)$ returns the sign of its input. The rank coefficient takes on

values between -1 and 1, where $\tau = 1$ denotes that the two rankings are the same and $\tau = -1$ denotes that the rankings are in reverse order of each other.

In this study, the value c is increased by increments of 0.1 until the rank coefficient between two subsequent rankings is greater than 0.99. The threshold of 0.99 is chosen somewhat arbitrarily, but one should generally choose a large threshold so that the rankings remain mostly the same (according to Kendall's rank coefficient) for the incrementally increasing bandwidths, making it appear less likely that an area has a large RDOS value by chance. As an additional note, in the extensive number of simulations that follow, the threshold of 0.99 is always passed by increasing the value c. Once the threshold is passed, the bandwidth is not increased further to avoid unnecessarily reducing the differences in the RDOS values, making it more difficult to split them into an inlier and an outlier set, as discussed next.

5.3.1.3 Choosing the number of observations to be identified as spatial outliers

Several methods can be used to choose the number of observations to be identified as spatial outliers from the ranked RDOS values, such as (1) using the top-n approach, where the observations with the n largest RDOS values are identified as outliers; (2) plotting the ranked RDOS values in decreasing order and identifying where they begin to level off (which looks like an elbow in the plot) to choose a point around which to split the observations into inliers and outliers; (3) using a clustering algorithm to split the RDOS values into an inlier and an outlier cluster. As a point of clarification, it should be noted that in this study, the inlier set contains any observation that is not identified as a singleton outlier. Hence, collective outliers are expected to be identified as part of the inlying data.

(1) The top-n approach is commonly used in the data mining literature (e.g., see Ramaswamy et al., 2000, Zhang et al., 2009, Aggarwal, 2017), where the researcher picks a fixed number of observations n or a proportion of the data to be identified as outliers. The method allows answering questions of the type "What are the ten most unusual observations in the data?". However, valuable information from the outlier scores is ignored when picking an arbitrary num-

ber of observations to be identified as outliers. Hence, other methods should be considered when identifying outliers from the RDOS values.

(2) A large RDOS value indicates that an area's observation is more dissimilar to its neighbouring observations than the neighbours are to their neighbouring observations. Hence, singleton outliers should have a noticeably larger RDOS value than the inlying data. Based on this concept, one can plot the RDOS values in decreasing order, with the RDOS value on the y-axis and its rank on the x-axis and identify where the values begin to level off. Visually, one can identify an elbow in the plotted values, similar to the elbow method used for within-sum-of-square plots in k-means clustering, which is attributed to Thorndike (1953) and explained in Section 2.8.1.3. One can pick a point near the elbow, possibly with a visual gap in the RDOS values, to split the observations into inliers and outliers around that point. Manually identifying the outlier and inlier sets is a subjective method, forcing the researcher to decide where to split the values into the two sets and possibly investigate the observations identified as outliers further to confirm that they appear to deviate from the typical local behaviour. A strength of the method is that it gives greater flexibility to the user and how they weigh the risks of making false positive and false negative assignments. For example, one researcher might prefer a smaller outlier set to reduce the risk of incorrectly identifying inliers as outliers. In comparison, another researcher might prefer a larger outlier set to reduce the risk of failing to detect some of the true outliers. The application in Section 5.5 gives a demonstration of this method.

(3) The simulation study in Section 5.4 evaluates the outlier detection method on many simulated datasets. While the manual assignment of observations to an inlier and an outlier set is the best motivated using the elbow method, it is not feasible to manually split the data into the two sets in each of these very large number of trials (e.g., there are 100 datasets for each of four simulation scenarios). Hence, the clustering method partitioning around medoids (PAM, recall Section 2.8.1.2) is used with two clusters to split the RDOS values into an

outlier and an inlier cluster to check if reasonable cutoffs in the RDOS values can be identified in this way.

5.3.2 A smoothing model that accounts for potential singleton outliers

In this study, a spatial smoothing model is proposed that accounts for the identified potential singleton spatial outliers. Section 5.3.2.1 describes the data likelihood and overall structure of the model, Section 5.3.2.2 describes the structure of the random effects, and Section 5.3.2.3 describes the prior specifications.

5.3.2.1 Data likelihood

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_K)$ denote observations from areas $(1, \ldots, K)$. Assume the spatial outlier detection method from Section 5.3.1 is applied to these observations (or some function thereof), and n_o areas are identified as outliers. Let $\mathcal{A}_{n_o}$ denote the set containing the n_o identified outliers, and $\mathcal{A}^c_{n_o}$ (the complement of $\mathcal{A}_{n_o}$) denote the set containing the $(K - n_o)$ identified inliers. Thus, $j \in \mathcal{A}_{n_o}$ if area j is identified as an outlier, and $j \in \mathcal{A}^c_{n_o}$, otherwise. Furthermore, let $\mathbb{1}_o(j)$ denote an indicator function so that $\mathbb{1}_o(j) = 1$ if $j \in \mathcal{A}_{n_o}$ (an outlier) and $\mathbb{1}_o(j) = 0$ if $j \in \mathcal{A}^c_{n_o}$ (an inlier). The proposed smoothing model takes the form

$$
\begin{aligned}
Y_j &\sim f(\mu_j), \\
g(\mu_j) &= \beta_0 + \boldsymbol{x}_j^T \boldsymbol{\beta} + (1 - \mathbb{1}_o(j))\phi_{j_c} + \mathbb{1}_o(j)\psi_{j_o}, \\
\boldsymbol{\psi} &\sim N\left(\mathbf{0}, \frac{1}{\kappa}\boldsymbol{I}\right) \in \mathbb{R}^{n_o}, \\
\boldsymbol{\phi} &= \frac{1}{\sqrt{\tau}}\left(\sqrt{1-\rho}\boldsymbol{v} + \sqrt{\rho}\boldsymbol{u}^*\right) \in \mathbb{R}^{K-n_o}, \\
\boldsymbol{v} &\sim N(\mathbf{0}, \boldsymbol{I}), \\
\boldsymbol{u}^* &\sim N(\mathbf{0}, \boldsymbol{Q}^*(\boldsymbol{W}^*)^-),
\end{aligned}
\tag{5.6}
$$

where each observation Y_j is modelled using a likelihood function $f(\mu_j)$ with mean μ_j. The model extends the existing BYM2 model (see Section 2.6.4) by the inclusion of separate random effects for the inlying and outlying observations. Since the outlying

observations appear to violate the spatial smoothness assumption, they are estimated independently from the inlying observations so that they do not contaminate the smooth risk (or prevalence) surface that is the target of inference for the inlying observations.

In disease mapping, the most common options are the Poisson and binomial likelihoods such that $Y_j \sim \text{Poisson}(\mu_j = e_j\theta_j)$, where e_j denotes an expected count computed via indirect standardisation and the relative risk θ_j is estimated using a log-linear model (i.e., of the form $\log(\theta_j) = \dots$), or $Y_j \sim \text{Binomial}(N_j, \alpha_j)$ with mean $\mu_j = N_j\alpha_j$, where N_j denotes the population size and the disease prevalence α_j is estimated using a logit-linear model (i.e., of the form $\log(\alpha_j/(1-\alpha_j)) = \dots$). For the general case with likelihood $f(\mu_j)$ from Equation (5.6), the transformed mean $g(\mu_j)$ is fitted as a linear model with intercept term β_0, covariates $\boldsymbol{x}_j = (x_{j1}, \dots, x_{jp})$ with regression coefficients $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$, and an added random effect that depends on whether area j was identified as a potential outlier.

For $j \in \mathcal{A}^c_{n_o}$ (inlier), the random effect associated with area j is ϕ_{j_c}, and the random effects vector of the $K - n_o$ areas identified as inliers is denoted $\boldsymbol{\phi} = \{\phi_j | j \in \mathcal{A}^c_{n_o}\}$. Similarly, for $j \in \mathcal{A}_{n_o}$ (outlier), the random effect associated with area j is ψ_{j_o}, and the random effects vector of the n_o areas identified as outliers is denoted $\boldsymbol{\psi} = \{\psi_j | j \in A_{n_o}\}$. Typically, disease risk and disease prevalence are assumed to change smoothly across the map (following Tobler's First Law of Geography). When no observations are identified as potential spatial outliers, Model (5.6) simplifies to a conventional spatial smoothing model that assigns a spatially correlated random effects vector $\boldsymbol{\phi}$ to the whole study region. When some areas are identified as outliers, their random effects $\boldsymbol{\psi}$ can be estimated independently from the correlated random effects $\boldsymbol{\phi}$, and the specific random effect structures applied in this study are described next.

5.3.2.2 Structure of the random effects

The modified smoothing model proposed in this study assigns independent random effects $\boldsymbol{\psi}$ to the observations that were identified as singleton outliers, and spatially correlated random effects $\boldsymbol{\phi}$ to the observations that were identified as inliers. The

outlying random effects $\boldsymbol{\psi}$ are assigned a zero-mean multivariate Gaussian prior distribution with precision matrix $\kappa \boldsymbol{I}$, where κ is a precision parameter and $\boldsymbol{I}$ denotes the identity matrix of size $n_o \times n_o$ (i.e., $\boldsymbol{\psi}$ is a vector of independent random effects). The inlier random effects $\boldsymbol{\phi}$ are assigned a BYM2 prior distribution (Simpson et al., 2017, recall Section 2.6.4). Specifically, $\boldsymbol{\phi} = \frac{1}{\sqrt{\tau}}\left(\sqrt{1-\rho}\boldsymbol{v} + \sqrt{\rho}\boldsymbol{u}^*\right)$, where τ is a precision parameter, $\boldsymbol{v}$ is a vector of independent random effects with a standard normal prior distribution, $\boldsymbol{u}^*$ is a vector of scaled correlated random effects with a joint intrinsic conditional autoregressive (ICAR, recall Section 2.6.1) prior distribution that enforces strong spatial autocorrelation, and ρ is a mixing parameter that controls the degree of spatial autocorrelation. For $\rho = 1$, the random effects $\boldsymbol{\phi}$ are estimated using only the ICAR prior distribution and for $\rho = 0$, they are estimated using only the independent Gaussian prior. For $0 < \rho < 1$, the BYM2 prior distribution is a mixture of the two.

The spatial structure of the ICAR prior distribution of the scaled random effects $\boldsymbol{u}^*$ is defined by the neighbourhood matrix $\boldsymbol{W}^*$ of size $(K - n_o) \times (K - n_o)$, where $w_{ij}^* = 1$ if inlying areas i and j are neighbours and $w_{ij}^* = 0$, otherwise. Generally, one could use a border-sharing rule so that two inlying areas are neighbours if they share a border. However, since the potential outliers are not included in the neighbourhood structure, there could be isolated inlying areas that share no border with other inlying areas. Hence, a k-nearest neighbour rule is applied so that $w_{ij}^* = 1$ if for inlying areas i and j, area j is among the k-nearest inlying areas of area i or area i is among the k-nearest inlying areas of area j, measured by the Euclidean distance between the areas' centroids. For the motivating study from Section 5.2, each area shares a border with an average number of 5.87 areas. Hence, in this study, the value is specified as $k = 6$ to approximately equal that number. The precision matrix $\boldsymbol{Q}^*(\boldsymbol{W}^*)$ is scaled so that the generalised variance $\sigma^2_{GV}(\boldsymbol{u}^*) = \exp\left(\frac{1}{K}\sum_{j=1}^{K}\log\left([\boldsymbol{Q}^*(\boldsymbol{W}^*)^-]_{jj}\right)\right)$ equals one. Scaling the precision matrix allows for interpretability of the hyperprior distribution of τ and for transferability between applications (Riebler et al., 2016).

As mentioned in Section 5.3.2.1, the modified smoothing model from Equation (5.6) smoothes over the observations that were identified as inliers but estimates the identified potential outliers independently. Since a BYM2 model is applied to smooth

over the identified inliers, the remainder of this chapter will refer to the modified smoothing model from Equation (5.6) as BYM2-O model, where the extension "-O" indicates that the modified model accounts for the potential outliers.

5.3.2.3 Prior specifications

The intercept β_0 is assigned a weakly informative normal prior distribution with mean zero and a variance of 1,000. Thus, the prior distribution imposes only a weak constraint and lets the data speak for themselves.

The random effects $\boldsymbol{\psi}$ of the observations identified as outliers are assigned a multivariate Gaussian prior distribution with mean zero, independent precision matrix and precision parameter κ. The log-precision is assigned a gamma hyperprior, so that $\log(\kappa) = \nu$ and

$$\pi(\nu) = \frac{b^a}{\Gamma(a)} \nu^{(a-1)} \exp(-b\nu), \tag{5.7}$$

where $a > 0$ is a shape parameter and $b > 0$ is the inverse-scale parameter. This study uses the `R-INLA` package for inference and here, the default settings $a = 1$ and $b = 0.00005$ are used.

The spatially correlated random effects $\boldsymbol{\phi}$ of the observations identified as inliers are assigned a BYM2 prior distribution with precision parameter τ and mixing parameter ρ. The parameters τ and ρ are assigned penalised complexity (PC) hyperprior distributions (recall Section 2.3.1.1). The precision parameter τ is assigned a type-2 Gumbel prior distribution of the form

$$\pi(\tau) = \frac{\theta}{2} \tau^{-3/2} \exp(-\theta \tau^{-1/2}). \tag{5.8}$$

Choosing the value of the parameter θ can be aided by considering the probability statement $P(1/\sqrt{\tau} > U) = \alpha$, resulting in $\theta = -\log(\alpha)/U$ (Riebler et al., 2016), where U and α can be chosen by the researcher. The default values in the `R-INLA` package are $U = 1$ and $\alpha = 0.01$, which are used in this study, since the corresponding probability statement reflects the prior assumption that there are no substantial spatial random effects, with a probability of 99% that the standard deviation takes

on a value smaller than one, so that the simplifying prior assumption is that the effects are close to their zero mean.

The mixing parameter ρ is assigned another PC prior distribution, where the base model is obtained for $\rho = 0$, i.e. the model where the independent component explains all variability of the random effects. The Kullback-Leibler divergence (KLD, e.g., recall Equation 2.7) for ρ between the base model and corresponding flexible BYM2 model is used to compute the distance scale $d(\rho)$, which is assigned an exponential prior distribution with parameter λ. In contrast to the precision parameter τ, the PC prior for ρ is not available in closed form, but can be computed in `R-INLA` for the specific neighbourhood matrix, and more details are provided in Riebler et al. (2016). The probability statement $P(\rho < U) = \alpha$ can be used to determine a reasonable value of λ, where the values of U and α can be chosen by the researcher. The default values in `R-INLA` are $U = 0.5$ and $\alpha = 0.5$, which are used in this study, since it is unknown whether or not to expect the random effects to be correlated before viewing the data.

5.4 Simulation study

Three simulation studies are conducted to answer the key questions about the proposed singleton outlier detection method and modified smoothing model: Are the outlier scores largest for true singleton outliers?; Is it possible to effectively identify a cutoff to split the outlier scores into inliers and outliers?; Does the modified model provide better estimates than a conventional smoothing model?

The geography used as the study region of these simulation studies is the local authority district (LAD) Liverpool, which is partitioned into $n = 298$ lower super output areas (LSOAs) and forms a subset of the study region of the asthma prevalence data presented in Section 5.2. Only a subset of the data is used for the simulation studies for computational reasons since the whole of England is too large to evaluate the methods on many generated datasets for different settings. The study investigates the proposed methods for discrete health outcomes, such as disease incidences or mortality counts. Section 5.4.1 presents the data generation framework and Section 5.4.2 reviews the outlier detection measures used for performance evaluation.

Section 5.4.3 presents the first simulation study, investigating if the outlier scores of the true singleton outliers rank in the top of the RDOS ranking and comparing the method's effectiveness to using local Moran's I for singleton outlier detection. Section 5.4.4 presents the second simulation study, investigating if it is possible to effectively identify a cutoff to split the RDOS values into inliers and outliers, where the number of observations to be identified as outliers is chosen using partitioning around medoids (PAM). Lastly, Section 5.4.5 compares the novel modified smoothing model that does not smooth over the identified outliers to a conventional smoothing model that smoothes over all areas.

5.4.1 Data generation

The data are generated for areas $j = 1, \ldots, K$ from the model

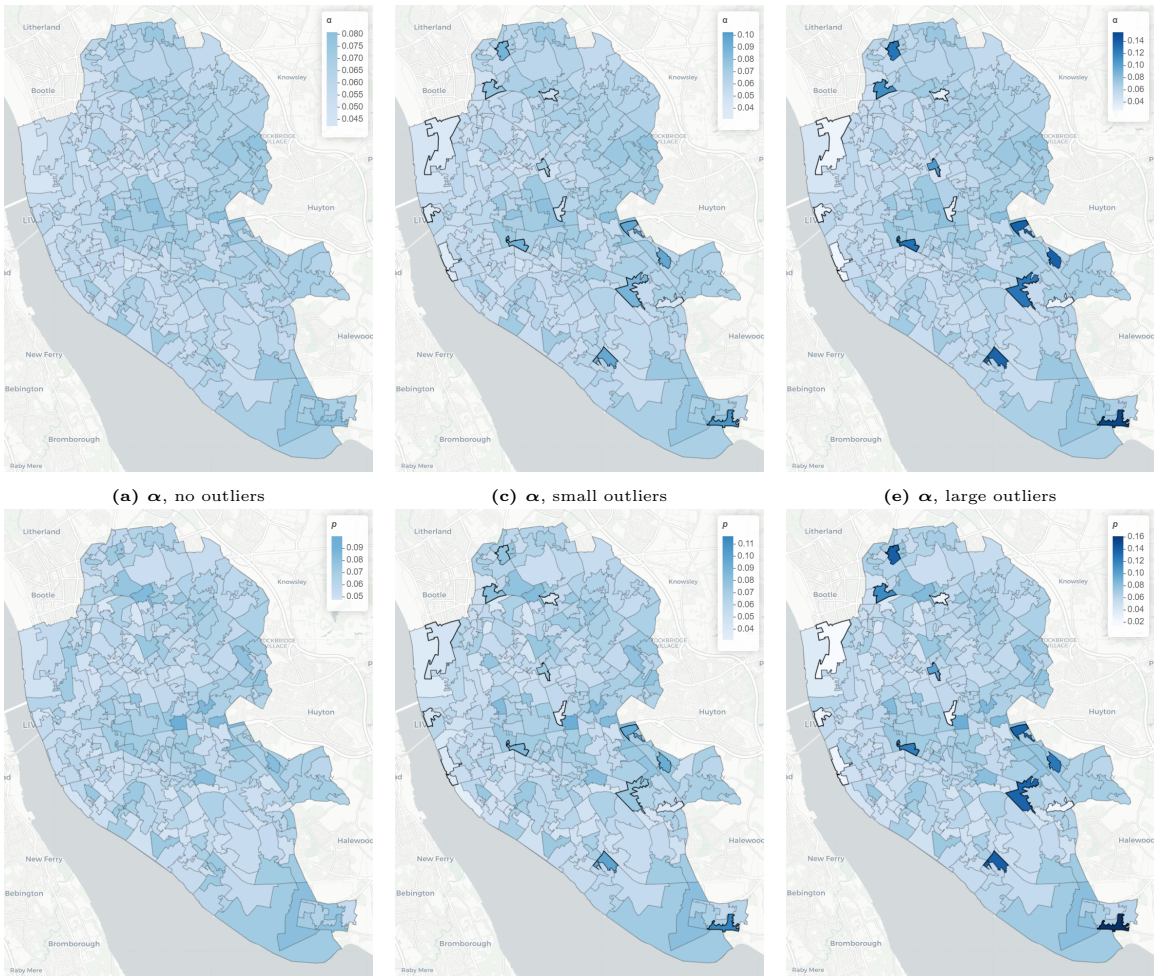
$$\begin{aligned} Y_j &\sim \text{Binomial}(N_j, \alpha_j), \\ \log\left(\frac{\alpha_j}{1-\alpha_j}\right) &= \beta_0 + \phi_j + \mathbb{1}_o(j)\gamma_j(\boldsymbol{\phi}) \times v, \\ \gamma_j(\boldsymbol{\phi}) &= \text{sgn}\left(\phi_j - \frac{1}{|\tilde{M}_j|}\sum_{i \in \tilde{M}_j} \phi_i\right), \\ \boldsymbol{\phi} &\sim N(\mathbf{0}, \tau^2 \boldsymbol{Q}(\rho, \boldsymbol{W})^{-1}). \end{aligned} \tag{5.9}$$

The observation Y_j is generated from a binomial likelihood, where the number of trials is the population size N_j of area j and α_j denotes the disease prevalence in that area, i.e., the true proportion of people in that area who have the disease. The observed prevalences $\boldsymbol{p} = (p_1, \ldots, p_K)$ are then computed as $p_j = Y_j/N_j$, for $j = 1, \ldots, K$. The population sizes are taken to be the numbers of registered patients from the asthma prevalence data presented in Section 5.2. The logit-transformed true prevalences are generated from a linear model with intercept β_0, spatial random effects $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)$, and a term $\gamma_j(\boldsymbol{\phi}) \times v$ that is added if area j is an outlier (i.e., if $\mathbb{1}_o(j) = 1$). The intercept term is set to the logit-transformed median of the observed prevalences from the asthma prevalence data in Section 5.2, as $\beta_0 = -2.73$.

The vector $\boldsymbol{\phi}$ consists of spatially correlated random effects generated from a Leroux CAR model (recall Section 2.6.2) with variance parameter τ^2 and precision

matrix $\boldsymbol{Q}(\rho, \boldsymbol{W}) = \rho(\text{diag}(\boldsymbol{W}\mathbf{1}) - \boldsymbol{W}) + (1-\rho)\boldsymbol{I}$, where ρ is a correlation parameter, $\boldsymbol{I}$ is a $K \times K$ identity matrix, $\mathbf{1}$ is a vector of ones of length K, and $\boldsymbol{W}$ denotes a symmetric neighbourhood matrix with $w_{ij} = 1$ if areas i and j are neighbours and $w_{ij} = 0$, otherwise. Here, $\boldsymbol{W}$ is a ten-nearest neighbour matrix based on the Euclidean distance between the areas' centroids, where additional neighbours are added so that $w_{ij} = 1$ if area j is among the ten-nearest neighbours of area i or area i is among the ten-nearest neighbours of area j and $w_{ij} = 0$, otherwise. The correlation parameter ρ determines the degree of spatial autocorrelation, and the random effects are generated for $\rho = 0.999$ to ensure they are spatially smooth. The variance parameter τ^2 determines the variability in the correlated random effects $\boldsymbol{\phi}$ and is set to $\tau^2 = 0.1$. The generated random effects are shifted so they sum to zero. The logit-transformed prevalence α_j equals the sum of the intercept term β_0 and spatial random effect ϕ_j if area j is an inlying area. For outlying areas, a shift parameter v is added so that the logit-transformed prevalence deviates more strongly from its neighbouring values than the usual deviation due to the variability in the random effects. The shift parameter's sign $\gamma_j(\boldsymbol{\phi})$ depends on the random effect values in area j and its neighbouring areas in the set $\tilde{M}_j = \{i : \text{ area } i \text{ shares a border with area } j\}$. The function $\text{sgn}(\cdot)$ returns the sign of its input: if ϕ_j is larger than the average of the random effects in the neighbourhood $\tilde{M}_j$, then $\gamma_j(\boldsymbol{\phi}) = 1$ and if ϕ_j is smaller than the average of the random effects in the neighbourhood $\tilde{M}_j$, then $\gamma_j(\boldsymbol{\phi}) = -1$. Therefore, each outlying logit-transformed prevalence is shifted by a magnitude v in the direction away from its neighbouring values, ensuring that it doesn't follow the typical local behaviour. Here, two scenarios are considered for the outlier magnitude v: small outliers with $v = 0.3$ and large outliers with $v = 0.7$. Furthermore, two scenarios are considered for the number of outliers in the data: $m = 3$ ($\sim 1\%$ of the data) and $m = 15$ ($\sim 5\%$).

Figure 5.3 shows example simulations for the data generated under Model (5.9) when no or $m = 15$ outliers are induced. Figure 5.3a shows a simulated prevalence vector $\boldsymbol{\alpha}$ when no outliers are present. Example prevalence vectors when 15 small or large outliers are induced are presented in Figures 5.3c and 5.3e, respectively, where the induced outliers are highlighted with a thicker border. Example maps



(a) $\boldsymbol{\alpha}$, no outliers (c) $\boldsymbol{\alpha}$, small outliers (e) $\boldsymbol{\alpha}$, large outliers

(b) $\boldsymbol{p}$, no outliers (d) $\boldsymbol{p}$, small outliers (f) $\boldsymbol{p}$, large outliers

Figure 5.3: Example maps showing simulated true prevalences $\boldsymbol{\alpha}$ and observed prevalences $\boldsymbol{p}$ when no or 15 outliers are induced.

of the observed prevalences $\boldsymbol{p}$ are presented in Figure 5.3b where no outliers are induced (corresponding to $\boldsymbol{\alpha}$ from Figure 5.3a), Figure 5.3d where 15 small outliers are induced (corresponding to Figure 5.3c), and Figure 5.3f where 15 large outliers are induced (corresponding to Figure 5.3e). For the larger outlier magnitude $v = 0.7$, the induced outliers can easily be identified from the observed prevalences $\boldsymbol{p}$. In contrast, for the smaller outlier magnitude $v = 0.3$, the observed prevalences of some of the induced outliers appear less unusual than those of some inlying observations, making it more difficult to distinguish between the outliers and some of the extreme inlying observations. Setting $v = 0.7$ allows evaluating the outlier detection method and modified smoothing model when clear singleton outliers are present, and $v = 0.3$ provides a more realistic scenario where the unusual true prevalence α_j for an outlying area j might not always be reflected in the corresponding observed prevalence p_j (e.g., see Figures 5.3c and 5.3d). The simulation studies in Sections 5.4.3, 5.4.4, and 5.4.5

are evaluated on 100 datasets that are generated under each of the four pairwise combinations of the number of outliers ($m = 3$ or $m = 15$) and the outlier magnitude (small or large outliers).

As a final note, when computing the RDOS values for these datasets, an initial bandwidth value is computed as a function of the data's local variability and increased by a multiplicative scalar c until the rankings of the RDOS values stabilise (recall Section 5.3.1.2). Table 5.1 shows the mean value and 95% uncertainty interval (in parentheses) of the scalar c over the 100 datasets for each of the four simulation settings. For the simulations with 15 small, three large, and three small outliers, the scalar c tends to take on a value between 2 and 3, with mean values of 2.20, 2.28, and 2.08 over the 100 datasets, respectively. For the scenario with 15 large outliers, the scalar c tends to take on slightly larger values, with a mean of 3.17 and a 95% uncertainty interval of (2.50,3.70).

Table 5.1: The mean and 95% uncertainty interval (in parentheses) of the scalar c over the 100 datasets for each of the four simulation settings.

	15 large outliers	15 small outliers	3 large outliers	3 small outliers
c	3.17 (2.50,3.70)	2.20 (2.00,2.50)	2.28 (1.90,2.80)	2.08 (1.80,2.30)

5.4.2 Outlier detection performance measures

This study evaluates the performance of the outlier detection method by computing its precision, sensitivity, and specificity (e.g., see Boehmke and Greenwell, 2019). Let $\mathcal{G}_m$ denote the set of the m outliers that are induced as described in Section 5.4.1, and let $\mathcal{A}_{n_o}$ denote the set of the n_o observations that are identified as outliers. The precision, sensitivity, and specificity are computed as

$$p_{\text{prec}} = \frac{|\mathcal{G}_m \cap \mathcal{A}_{n_o}|}{n_o}, \qquad p_{\text{sens}} = \frac{|\mathcal{G}_m \cap \mathcal{A}_{n_o}|}{m}, \qquad p_{\text{spec}} = \frac{|\mathcal{G}_m^c \cap \mathcal{A}_{n_o}^c|}{K - m}.$$

The set of genuine outliers $\mathcal{G}_m$ and its size m are unknown in real datasets. Hence, these measures can only be used in the simulation study in Section 5.4 but not for the motivating application in Section 5.5.

The precision is the proportion p_{prec} of the n_o identified outliers that are genuine outliers, where $|\mathcal{G}_m \cap \mathcal{A}_{n_o}|$ is the number of genuine outliers that are contained in the

set of observations that are identified as outliers. If the precision was used as the sole evaluation criterion, a small number of observations n_o could be identified as outliers to ensure that p_{prec} is large; i.e., if only the most unusual observations are identified as outliers, there is a higher chance that many of them will be genuine outliers. However, by choosing an n_o that is too small, a substantial number of genuine outliers might not be identified.

The sensitivity is the proportion p_{sens} of genuine outliers that are contained in the set of observations that are identified as outliers (i.e., the true positive rate). The sensitivity can be increased by selecting a larger number of observations n_o to be identified as outliers, increasing the chance that a greater proportion of the genuine outliers are included. However, one can maximise the sensitivity by setting $n_o = K$, i.e., identifying all observations as outliers results in $p_{\text{sens}} = 1$. Therefore, the goal is to choose n_o so that p_{prec} and p_{sens} are both large, with the optimal solution $p_{\text{prec}} = p_{\text{sens}} = 1$ obtained when $\mathcal{G}_m = \mathcal{A}_{n_o}$, i.e., when exactly the set of genuine outliers is identified as outliers.

As a third and final measure, this study also considers the specificity p_{spec}, which is the proportion of inliers that are not identified as outliers (i.e., the true negative rate). Here, $\mathcal{G}_m^c$ (the complement of $\mathcal{G}_m$) denotes the set of the $K - m$ inlying observations and $\mathcal{A}_{n_o}^c$ denotes the set of the $K - n_o$ observations that are not identified as outliers. The specificity p_{spec} is smallest when only inliers are identified as outliers so that the set $|\mathcal{G}_m^c \cap \mathcal{A}_{n_o}^c|$ contains $K - n_o - m$ observations. Thus, the specificity is bounded below by $(K - n_o - m)/(K - m) = 1 - n_o/(K - m)$, which will be close to one when $(K - m) >> n_o$. Generally, outliers are expected to make up only a small proportion of the data, so for application with very large K, the specificity will take on a large value as long as n_o is small.

In Section 5.4.3, receiver operator characteristic (ROC) curves are computed for different simulation settings to compare the RDOS-based outlier detection method to using local Moran's I statistic when applying the 'top-n' approach (where the observations with the n largest outlier scores are identified as outliers). The ROC curve plots the method's true positive rate (TPR, or sensitivity) against the false positive rate (FPR) as n_o (and hence, n in the top-n approach) is increased from 1

to K. The FPR is the complement of the specificity and computed as

$$\text{FPR} = \frac{|\mathcal{G}_m^c \cap \mathcal{A}_{n_o}|}{K - m}. \tag{5.10}$$

Thus, the false positive rate is the proportion of inliers incorrectly included in the set of observations identified as outliers. The ROC curve starts at FPR = TPR = 0 when no observations are identified as outliers ($n_o = 0$) and finishes at FPR = TPR = 1 when all observations are identified as outliers ($n_o = K$). The performance of different methods is compared by computing the area under the ROC curve (AUC), which takes on a maximum value of one when all genuine outliers are identified before incorrectly identifying any inlying observations as outliers, and a value of 0.5 corresponds to the expected result of a random classifier.

5.4.3 Simulation Study 1: Comparing the performance of RDOS and local Moran's I

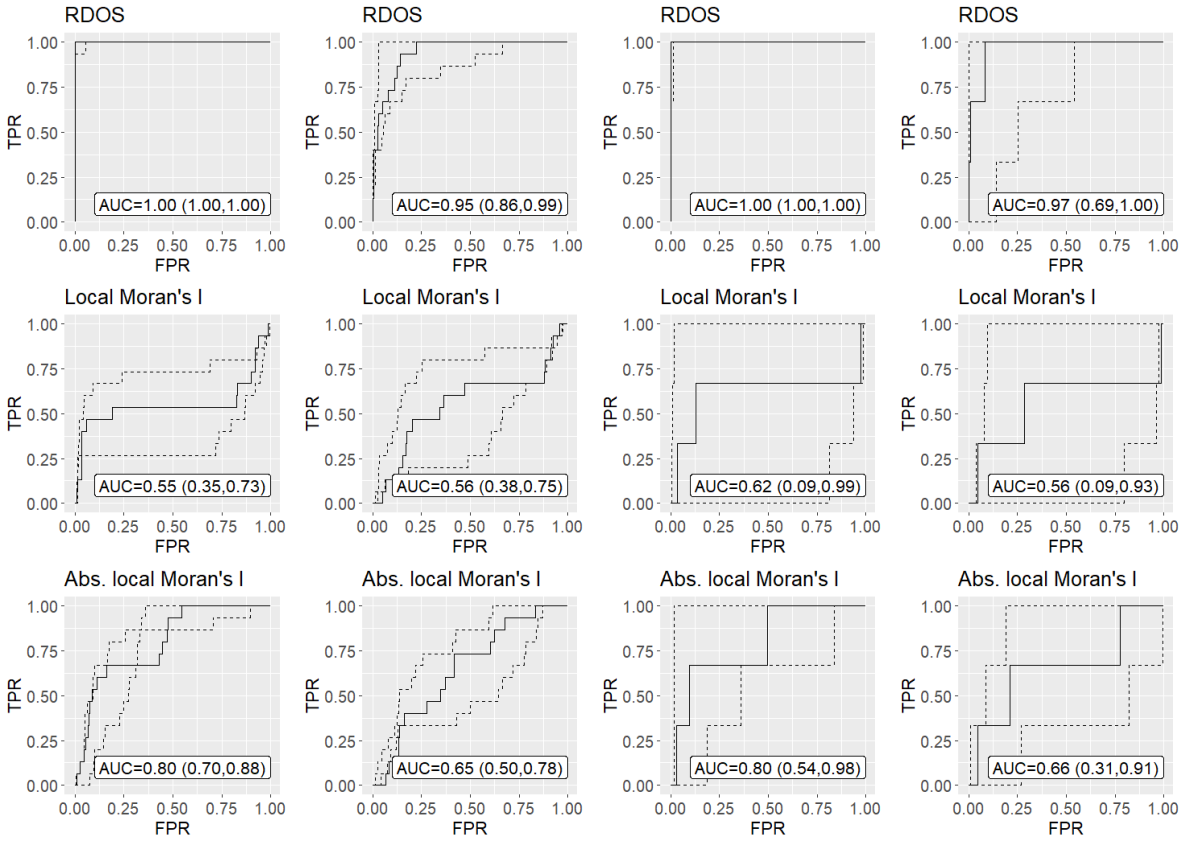
The local Moran's I statistic I_j (recall Section 2.5.1.2) takes on a negative value when y_j is either a high value surrounded by neighbouring values that tend to be low or a low value surrounded by neighbouring values that tend to be high, where high and low are in comparison to the overall mean $\bar{y}$. Hence, areas with negative local Moran's I values do not appear to follow the typical local behaviour and are suspected to be contextual or global outliers. Anselin (1995) claims that for the local Moran's I statistic, "a positive value indicates spatial clustering of similar values (either high or low)". However, it should be noted that singleton outliers could also result in a larger positive value, for example when an area has an unusually high value and is surrounded by neighbouring values that are on average slightly larger than the overall mean. Thus, when comparing the proposed method to using local Moran's I value for singleton spatial outlier detection, two options are considered: potential outliers are added to the outlier set either in order of increasing local Moran's I or decreasing absolute local Moran's I value.

This section compares the singleton spatial outlier detection performance when using the RDOS values to the existing method of using local Moran's I statistic (or

the alternative considering its absolute value) for areal data generated as described in Section 5.4.1. The neighbourhoods used for computing either measure are constructed using a k-nearest neighbour rule (according to the Euclidean distance between the areas' centroids) where additional areas are added to the neighbourhood sets so that area i is in the neighbourhood M_j if it is among the k-nearest neighbours of area j, or if area j is among the k-nearest neighbours of area i. The neighbourhoods should be small enough to be local but sufficiently large to accurately reflect the typical local behaviour. In this study, the number of nearest neighbours for constructing the neighbourhood sets is set to $k = 10$. Local Moran's I statistic is computed as in Equation (2.11), and the RDOS scores are computed as in Equation (5.4), with the bandwidth parameter value chosen as described in Section 5.3.1.2.

Figure 5.4 shows ROC curves computed for the three considered methods. The first method uses the RDOS values, where the observations are added to the outlier set by decreasing value. In the second method, observations are added to the outlier set by increasing local Moran's I value, and the third method ranks the observations by decreasing absolute local Moran's I value. The solid lines in the plots show the ROC curve corresponding to the median AUC (here, taken to be the 50th smallest AUC) and the dashed lines show the ROC curves corresponding to the 3rd smallest and 3rd largest AUC values and provide an approximate 95% uncertainty interval for the ROC curve under each scenario.

Most importantly, the plots show that using the RDOS values (top row) for singleton outlier detection delivers exceptional results, with median AUC values close to one even for the smaller outlier magnitude. For the larger outlier magnitude and either number of induced outliers (3 or 15), the ROC curves corresponding to the median AUCs appear perfect and the 95% uncertainty intervals of (1.00,1.00) confirm that for the RDOS values, the genuine outliers tend to be identified before any inlying observations. For the smaller outlier magnitude, some genuine outliers are not identified until after a considerable number of inliers have been incorrectly identified as outliers, resulting in AUC values of 0.95 (0.86,0.99) when $m = 15$ outliers are induced and 0.97 (0.69,1.00) when $m = 3$ outliers are induced. For example, for $m = 15$, the ROC curve corresponding to the median AUC reaches a true positive



(a) 15 large outliers **(b)** 15 small outliers **(c)** 3 large outliers **(d)** 3 small outliers

Figure 5.4: Median AUC and 95% uncertainty interval ROC curves showing the false positive rate (FPR) and true positive rate (TPR) when using the RDOS values or local Moran's I for outlier detection for data simulated under four simulation settings.

rate of one when the false positive rate is 0.2226, i.e., after 63 inliers have been incorrectly identified as outliers. Here, only three of the 15 induced outliers are identified before the first inlier is incorrectly identified as an outlier, and by the time eight of the induced outlier are identified, eight inliers have also been identified incorrectly. Thus, for the smaller outlier magnitude, choosing the set of observations to be identified as outliers from the ranked RDOS values comes as a trade-off between the false positive and false negative rate, which is further explored in Section 5.4.4. Given that the outlier magnitude $v = 0.3$ was intentionally chosen small so that some of the induced outliers do not appear unusual in the observed data (recall Figure 5.3d), the detection performance using the RDOS values is quite remarkable.

Another striking feature is that the singleton outliers are not generally associated with small local Moran's I values (middle row). As a result, adding the observations to the outlier set by increasing local Moran's I value is not a good method for singleton outlier detection, with only a fraction of the genuine outliers identified effectively. For $m = 15$ induced outliers, even when the outlier magnitude is large, the AUC

has a median value of only 0.55 with a 95% uncertainty interval of (0.35,0.73). As seen in the plot, many of the induced outliers are associated with some of the larger local Moran's I values and are among the last to be identified as outliers. The alternative ranking of the local Moran's I values by decreasing absolute value (bottom row) generally produces better detection results. For example, when $m = 15$ large outliers are induced, the AUC has a median of 0.80 with a 95% uncertainty interval of (0.70,0.88). Despite the overall improved detection performance, in the more realistic setting with a smaller outlier magnitude, the lower bound of the 95% uncertainty interval of the AUC is only 0.50 when $m = 15$ outliers are induced. For $m = 3$, the lower bound is even smaller, at 0.31. Thus, although the alternative ranking by decreasing absolute value produces overall better results than the conventional ranking by increasing value, the local Moran's I statistic is not generally a reliable measure to identify singleton outliers in areal data and using the proposed RDOS values appears to be a much more effective method.

5.4.4 Simulation Study 2: Using PAM to choose the outlier set from the RDOS values

This section evaluates the effectiveness of using the partitioning around medoids (PAM, recall Section 2.8.1.2) algorithm for splitting the RDOS values into inlier and outlier sets. The method is evaluated on the same 400 datasets as in Section 5.4.3, with 100 datasets for each of the four pairwise combinations of the number of outliers ($m = 3$ or $m = 15$) and the outlier magnitude (small or large outliers). Since it is not feasible to manually split the data into the two sets for all 400 datasets, the PAM algorithm with two clusters is used as an automated method. Table 5.2 presents the mean values and 95% uncertainty intervals (in parentheses) for the number of observations n_o that are identified as outliers and for the performance measures precision, sensitivity, and specificity over the 100 simulated datasets under each simulation setting.

The method is evaluated for different subsets of the RDOS values to be considered by the PAM algorithm. One option is to apply the PAM algorithm to all RDOS

Table 5.2: Simulation results for the outlier detection performance using partitioning around medoids (PAM) to split the RDOS values into an outlier and an inlier set.

		15 large outliers	15 small outliers	3 large outliers	3 small outliers
10% largest RDOS values	n_o	6.29 (2,10)	5.52 (1,12)	2.51 (1,7)	5.41 (1,13)
	p_{prec}	1.00 (1.00,1.00)	0.83 (0.38,1.00)	0.94 (0.43,1.00)	0.42 (0.00,1.00)
	p_{sens}	0.42 (0.13,0.67)	0.29 (0.07,0.57)	0.69 (0.33,1.00)	0.49 (0.00,1.00)
	p_{spec}	1.00 (1.00,1.00)	1.00 (0.99,1.00)	1.00 (0.99,1.00)	0.99 (0.96,1.00)
20% largest RDOS values	n_o	6.74 (2,12)	9.61 (1,22)	2.80 (1,14)	10.49 (1,24)
	p_{prec}	1.00 (1.00,1.00)	0.73 (0.39,1.00)	0.93 (0.22,1.00)	0.31 (0.00,1.00)
	p_{sens}	0.45 (0.17,0.80)	0.42 (0.07,0.70)	0.71 (0.33,1.00)	0.60 (0.00,1.00)
	p_{spec}	1.00 (1.00,1.00)	0.99 (0.95,1.00)	1.00 (0.96,1.00)	0.97 (0.93,1.00)
All RDOS values	n_o	7.17 (3,15)	46.60 (6,88)	26.08 (1,89)	63.93 (26,94)
	p_{prec}	1.00 (0.97,1.00)	0.34 (0.16,0.87)	0.63 (0.03,1.00)	0.05 (0.03,0.10)
	p_{sens}	0.48 (0.23,0.97)	0.81 (0.36,1.00)	0.85 (0.33,1.00)	0.94 (0.49,1.00)
	p_{spec}	1.00 (1.00,1.00)	0.88 (0.74,1.00)	0.92 (0.71,1.00)	0.79 (0.69,0.92)

values (in the bottom section of the table). However, one would generally expect the proportion of outliers in the dataset to be small (or zero) and hence, one could apply the PAM algorithm to a subset of the data to avoid the added difficulty in trying to detect small clusters (i.e., small sets of singleton outliers). Here, either $m = 3$ ($\sim 1\%$ of the data) or $m = 15$ ($\sim$5%) observations are induced outliers and hence, possible subsets should be larger than $m = 15$ to ensure these subsets contain at least some inliers. Hence, to investigate the variability in the detection results when using the PAM algorithm, the 30 ($\sim$ 10%, in the top section of the table) and 60 ($\sim$20%, in the middle section) largest RDOS values are considered as subsets to apply the PAM algorithm. It should be noted that smaller subsets could be considered, especially in applications where the proportion of genuine outliers is anticipated to be smaller than 5%.

The most important finding from Table 5.2 is that partitioning around medoids is not generally an effective method for splitting the RDOS values into an inlier and an outlier set. When 15 large outliers are induced in the dataset, fewer than half of them tend to be included in the outlier set, regardless of the initial number of RDOS values considered by the PAM algorithm (e.g., the average numbers are 6.29, 6.74, and 7.17 when applying PAM to the 10% largest, 20% largest, and all RDOS values, respectively). While only true outliers tend to be included in the outlier set (the lower bounds of the uncertainty intervals of the precision are 1.00, 1.00, and 0.97), the ROC curve from Section 5.4.3 suggests that most true outliers should rank highest

and hence, the sensitivity would be increased without sacrificing the high precision if some additional observations were identified as outliers. Hence, the detection results seem overly conservative when 15 large outliers are present.

For the scenario with 15 small outliers, the choice of the number of RDOS values to be considered by the PAM algorithm has a bigger influence on the results. Here, the average number of observations in the outlier set are 5.52 for the 10% largest RDOS values, 9.61 for the 20% largest RDOS values, and 46.60 when all RDOS values are considered. Thus, the outlier sets tend to be too large when all RDOS values are considered by the PAM algorithm. When only the 10% or 20% largest RDOS values are considered by the PAM algorithm, the outlier sets are smaller but tend to contain fewer than half of the induced outliers (the average sensitivity is 0.29 for 10% and 0.42 for 20%). However, in this scenario, increasing the size of the outlier sets might not lead to better splits, since some inliers tend to be already included erroneously (e.g., the average precision is 0.83 for 10% and 0.73 for 20%). These findings align with the results from Section 5.4.3, where the ROC curves for the simulations with 15 small outliers suggested that only half of the induced outliers tend to be identified before incorrectly identifying just as many (or more) inliers.

When only three outliers are induced in the dataset, the outlier sets tend to be much too large for either outlier size (large or small) when all RDOS values are considered by the PAM algorithm (e.g., on average, 26.08 and 63.93 observations are included, respectively). The size of the outlier sets tends to be much reduced when applying the PAM algorithm to only the 10% or 20% largest RDOS values. For example, for the larger outliers, the average number of observations n_o included in the outlier set are 2.51 for the 10% largest RDOS values and 2.80 for the 20% largest RDOS values, and most observations included tend to be true outliers (e.g., the average precision is 0.94 and 0.93, respectively). For the smaller outliers, the outlier sets tend to be larger, with average sizes of 5.41 and 10.49 for the 10% and 20% largest RDOS values, respectively, and the precision tends to be substantially lower (e.g., the average precisions are 0.42 and 0.31).

Generally, for the two scenarios with the smaller outliers and when only a subset of RDOS values are considered by the PAM algorithm, the sizes of the resulting outlier

sets tend to be fairly similar regardless of the true number of outliers induced. For example, when the 10% largest RDOS values are considered by them PAM algorithm, the outlier sets have an average size of 5.52 (1,12) when 15 small outliers are induced and 5.41 (1,13) when only three small outliers are induced. Thus, the PAM algorithm does not appear to be a reliable method for splitting the RDOS values into an inlier and an outlier set, especially when the outlier magnitude is small.

Since the PAM algorithm is not generally an effective method for splitting the RDOS values into the two sets, for real applications, it is advised to plot the RDOS values in decreasing order to choose a cutoff manually (revisit Section 5.3.1.3 for further explanations) and investigate the observations near the cutoff between the so obtained outlier and inlier sets further to check if the outliers appear indeed more unusual than the inlying observations. The manual approach will be demonstrated in Section 5.5.

5.4.5 Simulation Study 3: Evaluating the modified smoothing model

As a final step of the proposed method, the modified smoothing model BYM2-O from Equation (5.6) in Section 5.3.2 is applied to the same 400 simulated datasets that were already analysed in Sections 5.4.3 and 5.4.4, and its accuracy is compared to that of a conventional BYM2 model. Specifically, the study considers the following models:

- The conventional BYM2 model.
- The modified smoothing model with outlier sets obtained from the PAM algorithm
  - considering the 10% largest RDOS values, referred to as "BYM2-O, PAM (10%)".
  - considering the 20% largest RDOS values, referred to as "BYM2-O, PAM (20%)".
  - considering all RDOS values, referred to as "BYM2-O, PAM (all)".

- The modified smoothing model with the outlier set equal to the set of induced outliers, referred to as "BYM2-O, best case".

For the models "BYM2-O, PAM (10%)", "BYM2-O, PAM (20%)", and "BYM2-O, PAM (all)", the evaluations pretend the outliers are unknown and the BYM2-O model from Equation (5.6) accounts for the potential singleton spatial outliers identified from the RDOS values and the PAM algorithm as presented in Section 5.4.4. The likelihood in Model (5.6) is $Y_j \sim \text{Binomial}(N_j, \alpha_j)$ with number of trials N_j equal to the number of registered patients from the asthma prevalence data, and the generated (but pretended to be unknown) prevalence α_j is the target for inference, for $j = 1, \ldots, K$, and fitted using a logit-linear model. The prevalence is inferred from the generated disease counts using the modified model BYM2-O, and the conventional BYM2 model is chosen for comparison. Section 5.4.4 showed that the detection results using the PAM algorithm were not optimal for most simulation settings. Hence, additionally to the outlier sets identified in Section 5.4.4, the BYM2-O model is fitted under a "best case" scenario referred to as "BYM2-O, best case", where exactly the true outliers are included in the identified outlier set to check how close the modified smoothing model under sub-optimal outlier identification can get to this optimal scenario.

In these simulations, the true disease prevalence is known and so each model's efficacy can be evaluated by how well the fitted values estimate the true values. Another consideration is how much variability there is in the fitted values. Specifically, the following analysis uses the root mean square error (RMSE) and median absolute error (MAE) to report the overall differences between the true disease prevalence and the posterior mean fitted values, and reports the coverage (the proportion of times the true prevalence is contained in the 95% credible interval) and the mean credible interval width as measures of variability in the fitted values. These measures are computed over all observations for each of the four simulation settings: 15 ($\sim 5\%$) large outliers, 15 small outliers, 3 ($\sim 1\%$) large outliers, and 3 small outliers. It should be noted that the results for the BYM2-O models that use the outlier sets obtained from the PAM algorithm applied to all observations are presented for completeness. However, Section 5.4.4 showed that the PAM algorithm tends to produce

outlier sets that are much too large when applied to all RDOS values and hence, the results for the fitted models tend to be much worse than for the scenarios where the PAM algorithm is applied to only a subset of the RDOS values. Hence, the following analysis focuses on the results obtained for the BYM2-O models that use the outlier sets obtained from the PAM algorithm applied to the 10% or 20% largest RDOS values.

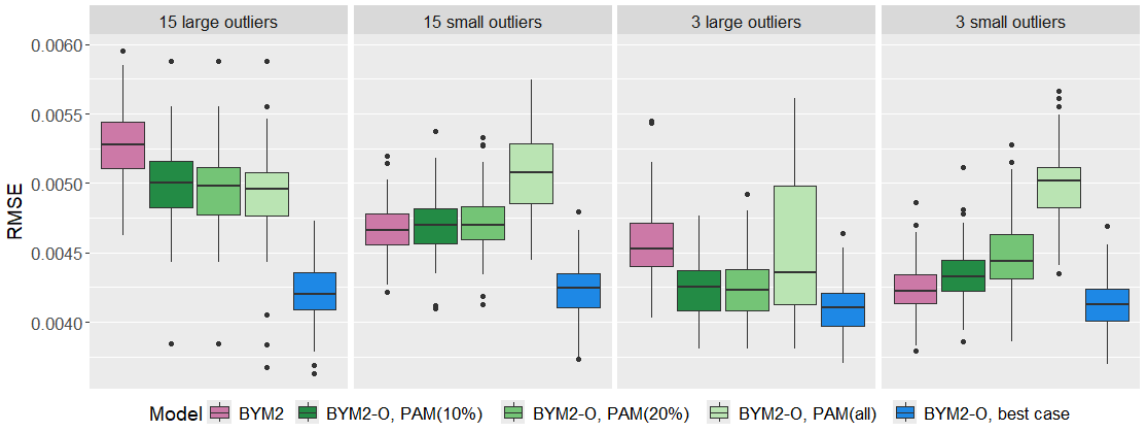


Figure 5.5: RMSE for the five considered models under the four simulation settings.

Figure 5.5 shows boxplots summarising the RMSE for each of the five models (the conventional BYM2 model, three BYM2-O models for outlier sets identified with PAM, and the "best case" BYM2-O model) for the four simulation scenarios, each evaluated over 100 simulated datasets. When the outliers are large, the RMSE over all observations is substantially lower for the modified smoothing models than the conventional BYM2 model. The overall reduced RMSE can be seen for both scenarios with 15 and 3 large outliers. For the scenario with 15 large outliers, the RMSE for the "best case" scenario BYM2-O model tends to be much lower than for the BYM2-O models that use the outlier sets identified using the PAM algorithm (e.g., the average RMSE is 0.0042 for the "best case" scenario, and 0.0050, 0.0047, and 0.0049 for the BYM2-O models with outlier sets identified using the PAM algorithm applied to the 10% largest, 20% largest, and all RDOS values, respectively). Recall from Section 5.4.4 that in this scenario, the PAM algorithm tends to identify only just under half of the induced large outliers. Hence, in this scenario, more than half of the induced outliers will tend to still be included in the inlier set, which explains the substantially larger RMSE values than in the "best case" scenario. Nonetheless, even

for the sub-optimal outlier identification, the BYM2-O models provide overall better point estimates than the conventional BYM2 model (which has an average RMSE of 0.0053 in the scenario with 15 large outliers), which suggests that the conventional smoothing model can be improved by estimating potential outliers independently from the smooth prevalence surface when some of the outliers are large.

When the outliers are small, the RMSE tends to be slightly larger for the BYM2-O models than the conventional BYM2 model when the outlier sets are identified using the PAM algorithm. For the scenario with 3 small outliers, even for the "best case" scenario, the RMSE is only slightly lower than for the conventional BYM2 model. For the scenario with 15 small outliers, the RMSE under the "best case" scenario tends to be substantially lower than for the conventional BYM2 model, which suggests that identifying the outliers correctly would generally lead to much better point estimates. However, for the outlier sets obtained from the PAM algorithm, too many observations tend to be incorrectly identified (inliers are contained in the outlier set and outliers are not identified) and as a result, the RMSE is slightly larger than for the conventional BYM2 model.

It should be noted that the RMSE is a non-robust summary statistic of the differences between the true and the fitted prevalence, as its value could be inflated by a few large errors. Thus, a few observations that are badly estimated could have a large impact on the RMSE. Instead, one can consider the median absolute error (MAE), which provides a more robust estimate of how similar the point estimates are to the true values, overall. Figure 5.6 shows boxplots that summarise the MAE values of the five models over the 100 datasets under each of the four simulation scenarios. The figure confirms that the point estimates of the BYM2-O models tend to be better than those of the conventional BYM2 model when the outliers are large. Additionally, the figure reveals that the BYM2-O model tends to also provide overall better point estimates in the scenario with 15 small outliers. Thus, while the point estimates of the BYM2-O model might be worse for some observations (resulting in a tendency of observing a slightly larger RMSE, as seen in Figure 5.5), they appear to be overall closer to the true prevalence than those obtained from the conventional BYM2 model. In contrast, when only three small outliers are present, the MAE tends

to be slightly larger for the BYM2-O models than for the conventional BYM2 model. This is not entirely surprising, given that the few and small outliers are unlikely to have a strong impact on the estimation of the smooth risk surface. Thus, applying a modified smoothing model appears warranted only if a larger number of outliers or (at least) a few outliers with a larger outlier magnitude are present in the data.

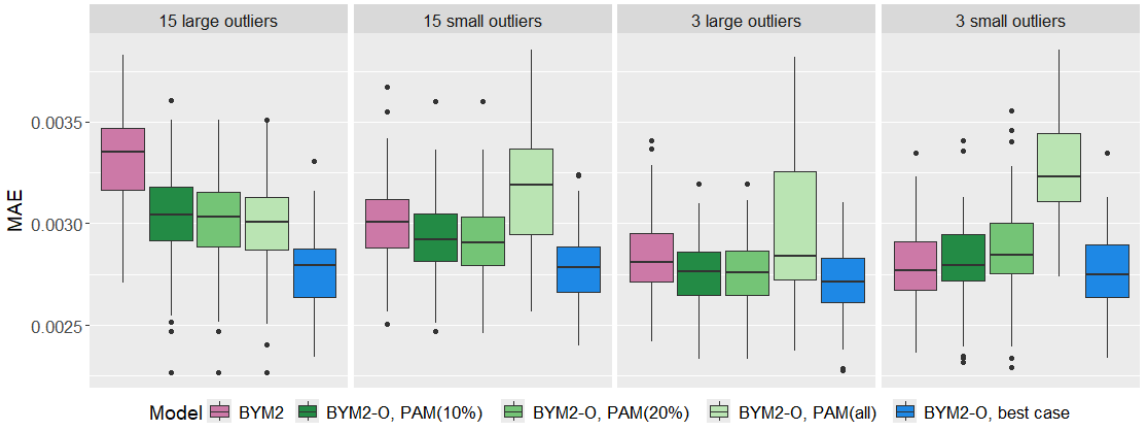


Figure 5.6: Median absolute error (MAE) for the five considered models under the four simulation settings.

Another consideration for the evaluation of the model fitting is the variability in the fitted prevalence. Figure 5.7 shows boxplots summarising the coverage (the proportion of times the true prevalence is contained in the 95% credible interval) of the five models over the 100 simulated datasets in each of the four simulation scenarios. Here, the optimal value of 0.95 is indicated by the orange horizontal line.

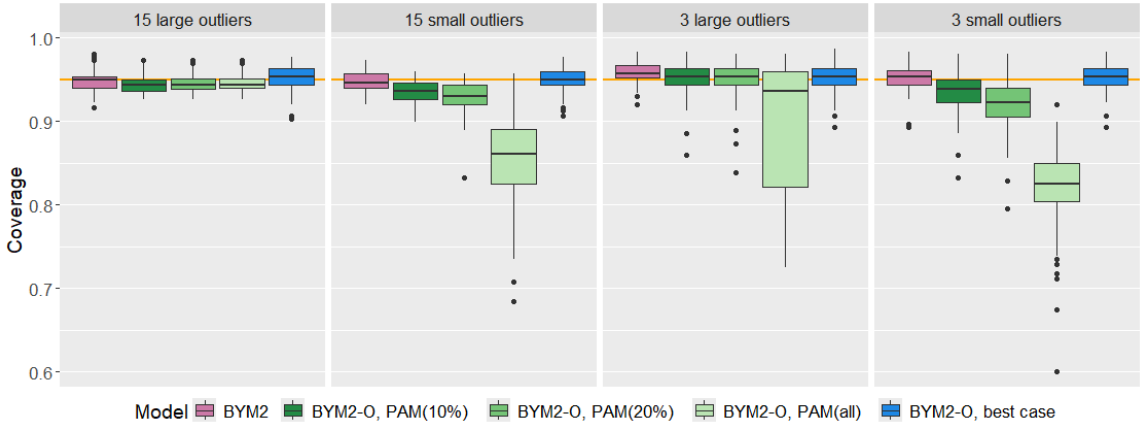


Figure 5.7: Coverage for the five considered models under the four simulation settings.

The figure shows that the coverage tends to be close to 0.95 for the conventional BYM2 model and the BYM2-O models using the outlier sets obtained from the PAM algorithm that considers the 10% or 20% largest RDOS values. However, when

the outlier magnitude is small, the BYM2-0 models tend to have a slightly lower coverage (with median values between 0.92 and 0.94). The modified model under the best case scenario tends to have a coverage of approximately 0.95. Hence, the other smoothing models likely have a slightly lower coverage due to the incorrectly identified observations from the inlier and outlier sets. The coverage is likely impacted by the credible interval widths, which are presented in Figure 5.8.

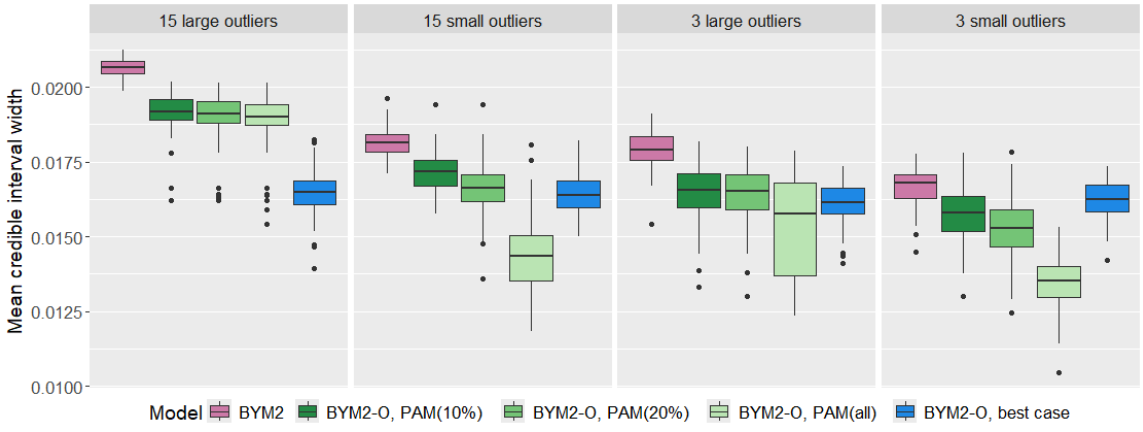


Figure 5.8: Mean credible interval width for the five considered models under the four simulation settings.

Figure 5.8 shows boxplots summarising the mean credible interval width for the five models over the 100 simulated datasets for each simulation scenario. The figure shows that the average credible interval width tends to be substantially smaller for the modified smoothing models than for the conventional BYM2 model. For the scenario with 15 large outliers, the average credible interval width of the "best case" BYM2-O model tends to be substantially smaller than for the other modified smoothing models, again likely because when the PAM algorithm is used, only just under half of the induced outliers tend to be identified. For the other simulation settings, the mean credible interval width of the BYM2-O models that use the outlier sets obtained using PAM tend to be much closer to that of the "best case" BYM2-O model and for the scenario with three small outliers, they tend to be even below that of the "best case" BYM2-O model. For the scenarios with 15 induced outliers (either large or small) and 3 large outliers, the median absolute errors tend to be smaller for the BYM2-O models (recall Figure 5.6) and the coverage tends to be close to 0.95 (recall Figure 5.7). Hence, the smaller mean credible interval widths lead to more precise estimates. In contrast, for the scenario with 3 small outliers, the median absolute

errors tend to be larger for the BYM2-O models than the conventional BYM2 model and hence, the smaller mean credible interval widths will tend to lead to more biased estimates.

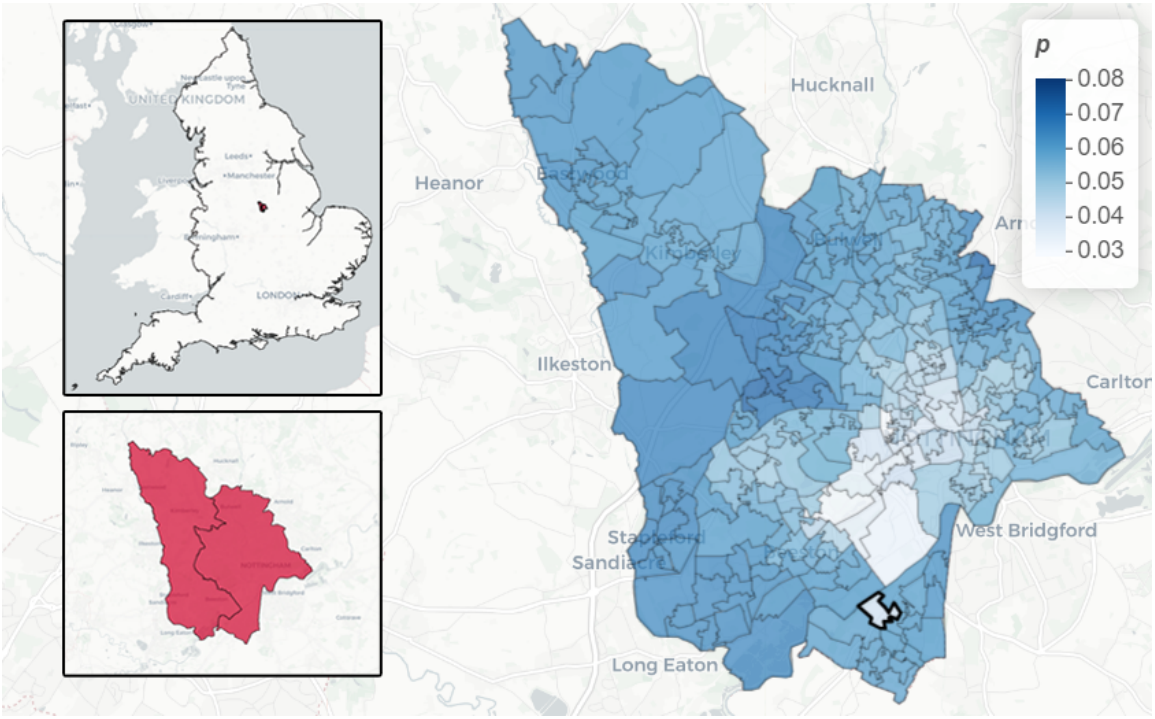
As a conclusion for the model fitting evaluation, the BYM2-O models appear to provide better estimates than the conventional BYM2 model, but only when at least some of the outliers are large or a larger number of outliers are present. Under those scenarios, the modified smoothing models tend to provide overall better prevalence estimates than the conventional BYM2 model. For the scenarios with small outliers, the BYM2-O models might not lead to improved fitted values or even produce biased estimates, which highlights the importance of visually exploring the identified outliers to decide if a modified model should be applied.

5.5 Application: Asthma prevalence at a small area level in England

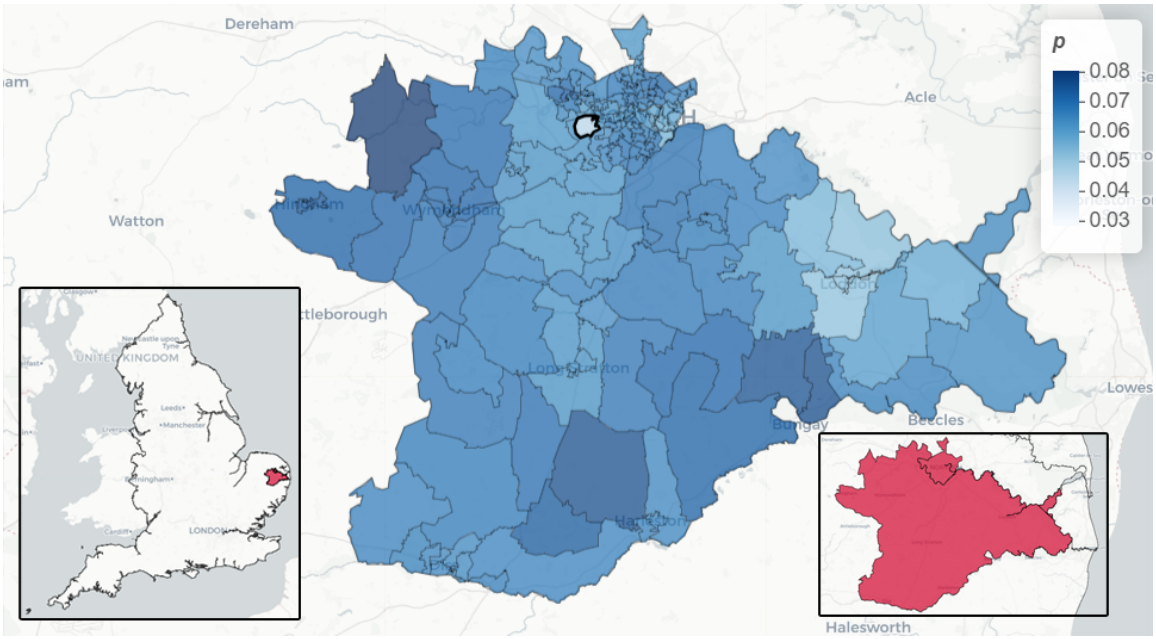
Here, the proposed methodology is applied to the asthma prevalence data presented in Section 5.2. Section 5.5.1 considers possible options for splitting the observations into inlier and outlier sets and proposes a split for the further analysis, answering the first motivating question, "Which areas appear to be potential outliers?". Section 5.5.2 presents an exploratory analysis of the identified outlier set to answer the second motivating question, "Do the identified outliers appear to be contextual or global?". Lastly, Section 5.5.3 presents the results from fitting the identified inlier/outlier sets using the BYM2-O model and shows how these compare to the results obtained from a conventional BYM2 model to answer the third motivating question, "How do the potential outliers affect the modelling?".

5.5.1 Which areas appear to be potential outliers?

The RDOS values are computed for the observed asthma prevalences, over neighbourhoods that are created using the k-nearest neighbour rule and made symmetric so that if area i is among the k nearest neighbours of area j, then area j is added to



(a) The area with the largest RDOS value.



(b) The area with the 2nd largest RDOS value.

Figure 5.9: Maps showing the observed asthma prevalence for the local authority districts containing or bordering the areas with the largest and second largest RDOS values.

the set of neighbours of area i. Here, the value $k = 10$ is used again, since it seems sufficiently small to compare the areas to their nearby areas and sufficiently large to capture the local behaviour (the value $k = 10$ was used in the simulation studies, with exceptional detection results presented in Section 5.4.3). The bandwidth parameter h_c is computed as $h_c = c \times \underset{j=1,\ldots,K}{\text{median}}\left(\underset{i \in M_j}{\text{median}}(|p_i - p_j|)\right)$, where one has to specify the value of the scalar c and p_j denotes the observed asthma prevalence, for

$j = 1, \ldots, K$. The smallest value of c for which Kendall's rank coefficient is greater than 0.99 (which indicates that the rankings stabilise; recall Section 5.3.1.2) is $c = 3.2$ and the bandwidth h_c is computed for that value.

Generally, one can plot the RDOS values to split the observations into inliers and outliers. However, this should only be done when at least some of the observations appear to be outliers. Hence, before looking at the distribution of all RDOS values, maps of some observations with the largest RDOS values are consider to see if these observations appear unusual. Figure 5.9 presents maps that show the observed asthma prevalences in the areas with the two largest RDOS values (these areas are highlighted with a thicker border). Note that the same colour scale is used for the two maps, so one can compare the observed asthma prevalence between these maps. Figure 5.9a shows the area with the largest RDOS value, which has a relatively low observed asthma prevalence of 0.0409 (the median observed prevalence is 0.0611) and is surrounded by areas with observed prevalences that are slightly larger (e.g., the border-sharing areas have observed asthma prevalences ranging from 0.0602 to 0.0624). The area is located in the East Midlands, and the upper boxed image in Figure 5.9a shows the location of the LSOAs from the larger map on the map of England, highlighted in red. More specifically, the outlying area is located in the local authority district (LAD) Nottingham on the border to the LAD Broxtowe, and the two LADs are shown in the lower boxed image, with Nottingham being the LAD further East. The area with the second largest RDOS value (presented in Figures 5.9b) also has a relatively low observed asthma prevalence of 0.0455 and is surrounded by areas with slightly larger observed asthma prevalences (e.g., the border-sharing areas have asthma prevalences ranging from 0.0601 to 0.0655). The outlying area is located in the LAD Norwich (East of England), on the border to South Norfolk and hence, Figure 5.9b shows the observed asthma prevalence for Norwich and South Norfolk. The lower boxed image shows the two LADs, where Norwich is the LAD further North. The areas with the two largest RDOS values highlighted in Figure 5.9 both appear to deviate from their typical local behaviour and hence, they can be identified as potential outliers. The question remains if these outliers are contextual or global, which is further investigated in Section 5.5.2. Since at least some potential

outliers appear to be present in the data, the next step is to choose the number of observations to be identified as outliers.

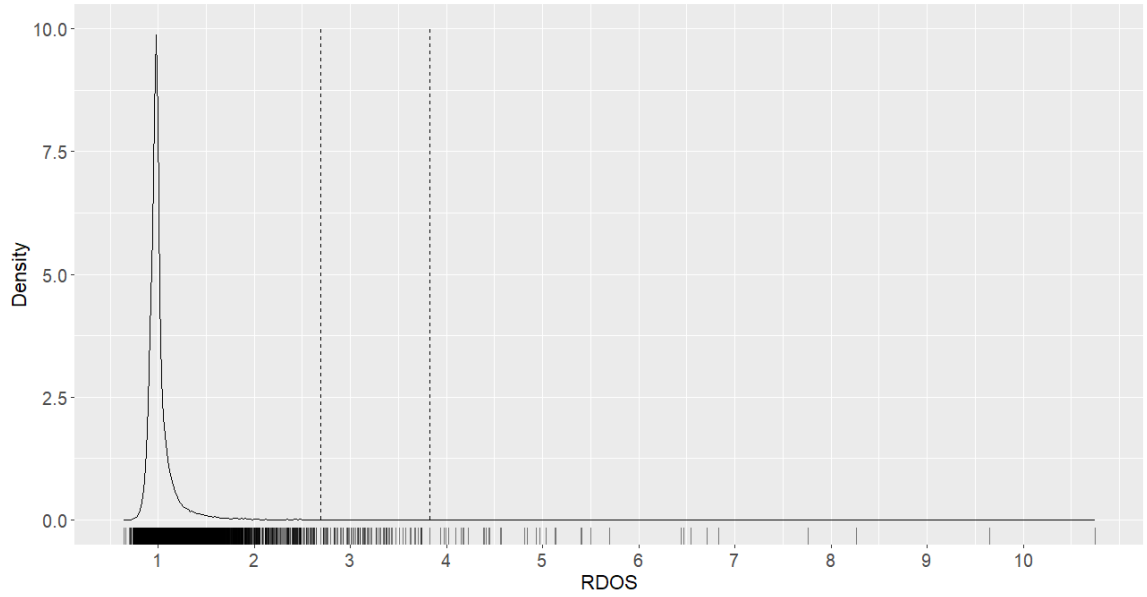
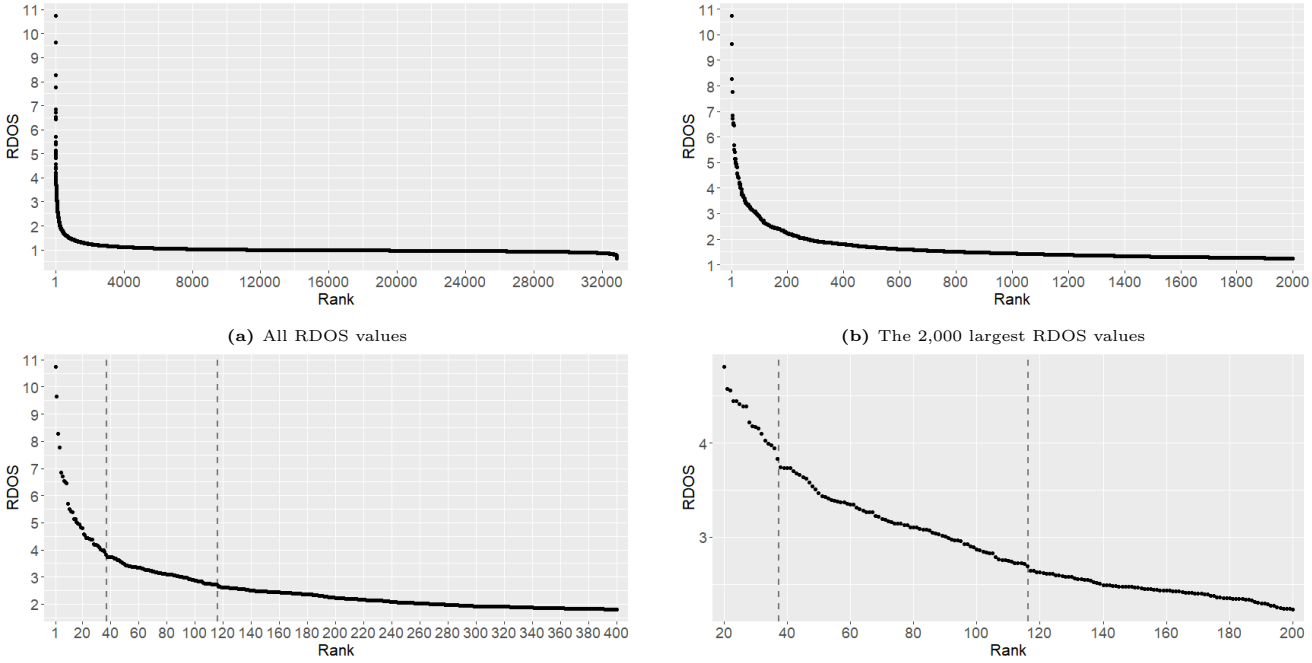


Figure 5.10: Density rug plot over all RDOS values with two possible inlier/outlier splits indicated by the dashed vertical lines.

Figure 5.10 shows a rug density-plot over all 32,844 RDOS values, where each vertical line below the x-axis represents an observed RDOS value and the estimated density curve is shown in the plot. The plot can be used to identify larger gaps in the RDOS values to split them into inliers and outliers, since the outliers are anticipated to have larger RDOS values than the inliers. Two possible inlier/outlier splits are indicated by the vertical dashed lines. One can increase the number of observations to be included in the identified outlier set by moving the dashed vertical line from right to left. The first split considered in Figure 5.10 is between the 37th and 38th largest RDOS value (indicated by the right dashed vertical line), as there appears to be a large gap and the 37 largest RDOS values look to be more sparsely distributed than the RDOS values to the left of that split. The second possible split is between the 116th and 117th largest RDOS values (indicated by the left dashed vertical line), where a final larger gap is noticeable when moving along the x-axis from right to left. Thus, the two considered outlier sets contain 37 and 116 potential outliers, respectively.

Additionally to the rug density plot, one can consider plots of the ranked RDOS values in decreasing value to look for an 'elbow' in the plot, as described in Sec-



(a) All RDOS values

(b) The 2,000 largest RDOS values

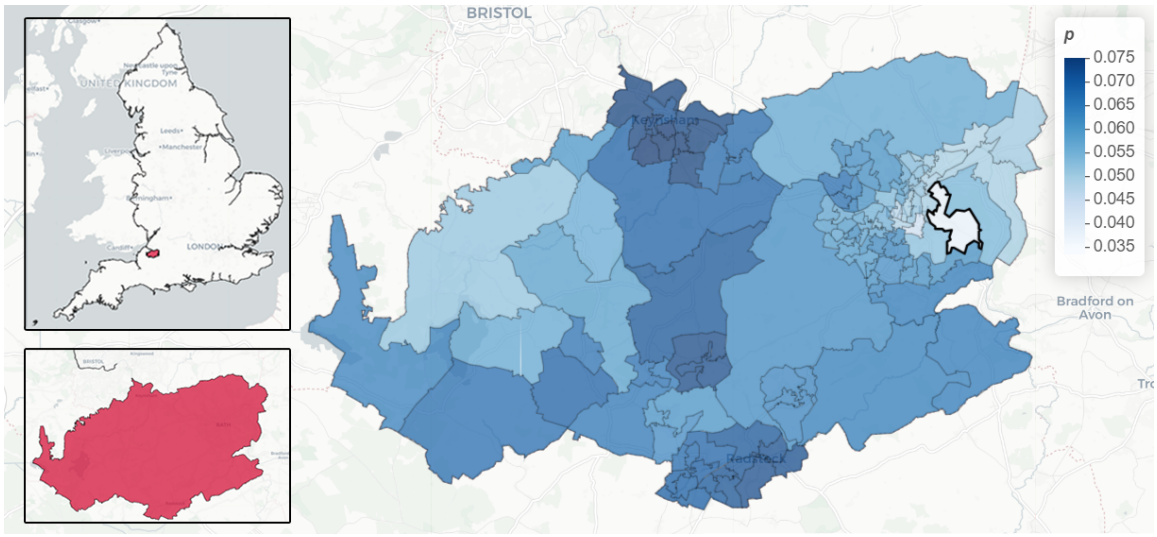
(c) The 400 largest RDOS values

(d) The 20th to 200th largest RDOS values

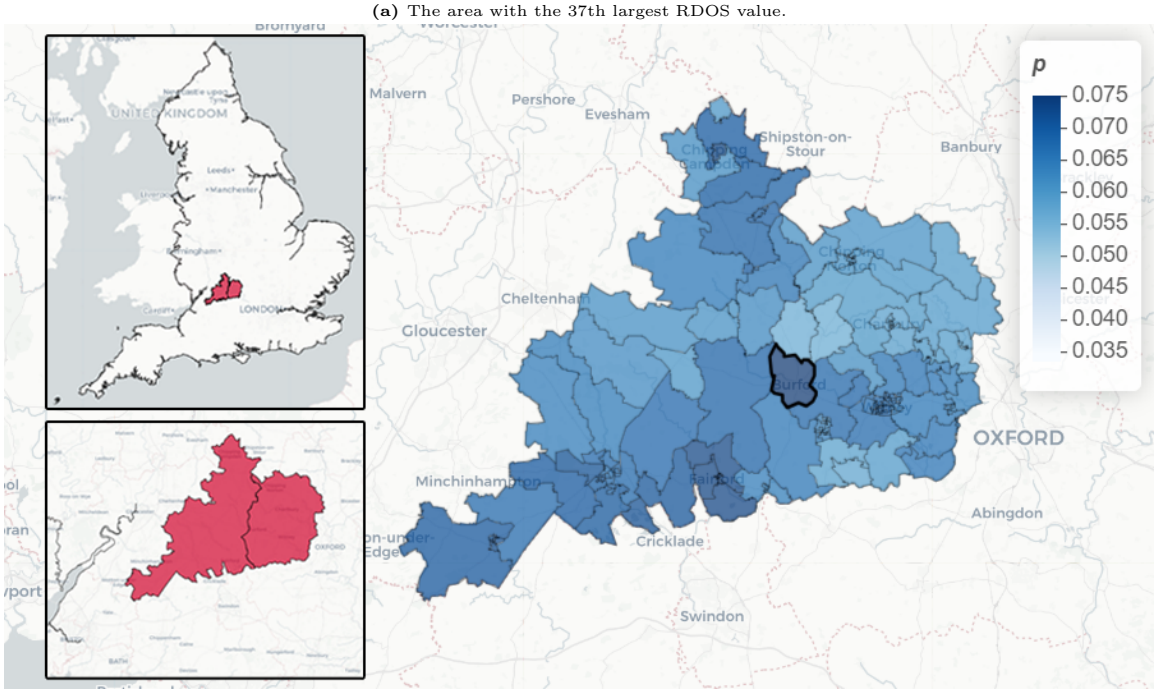
Figure 5.11: RDOS values for the asthma prevalence data in decreasing order.

tion 5.3.1.3 and presented in Figure 5.11. Figure 5.11a shows the RDOS values of all 32,844 areas. Moving from left to right, the RDOS values begin to level off somewhere among the 2,000 largest values. Figure 5.11b shows the 2,000 largest RDOS values, where some of the 400 largest RDOS values appear substantially larger than the rest. Figure 5.11c shows only the 400 largest RDOS values. Here, the same inlier/outlier splits from Figure 5.10 are represented by the dashed vertical lines. The gaps identified in Figure 5.10 are less noticeable in Figure 5.11c since many of the points tend to overlap, due to the large number of observations presented in the plot. Figure 5.11d shows only the 20th to the 200th largest RDOS values, where the gaps identified in Figure 5.10 can be seen more easily than in the other plots in Figure 5.11.

One should check manually which observations appear to be potential outliers, which can be done by visualising the observations on a map. Figure 5.12 shows the areas with the 37th and 38th largest RDOS values, highlighted with a thicker border. Again, the colour scale is the same for these two maps. Figure 5.12a shows the area with the 37th largest RDOS value, which is the area with the smallest RDOS value in the outlier set of size 37. The area has an observed asthma prevalence of 0.0381. Although the area is in a region with other areas that have smaller observed asthma prevalences, its observed asthma prevalence is noticeably lower (e.g., its bordering



(a) The area with the 37th largest RDOS value.



(b) The area with the 38th largest RDOS value.

Figure 5.12: Maps showing the observed asthma prevalence for the local authority districts containing or bordering the areas with the 37th and 38th largest RDOS values.

areas have observed asthma prevalences between 0.0493 and 0.0625) and hence, one could argue that it should be identified as a singleton outlier. Figure 5.12b shows the area with the 38th largest RDOS value. For the smaller outlier set of size 37, this area would be the inlying area with the largest RDOS value. Although the area's observed asthma prevalence of 0.0759 is above the median of 0.0611 and the areas at its Northern border have lower observed asthma prevalences of 0.0563 and 0.0566, its other neighbouring areas tend to have also slightly larger observed asthma prevalences (the other values are 0.0665, 0.0690, and 0.0697) and hence, the area's observed asthma prevalence does not appear to deviate substantially from the typical

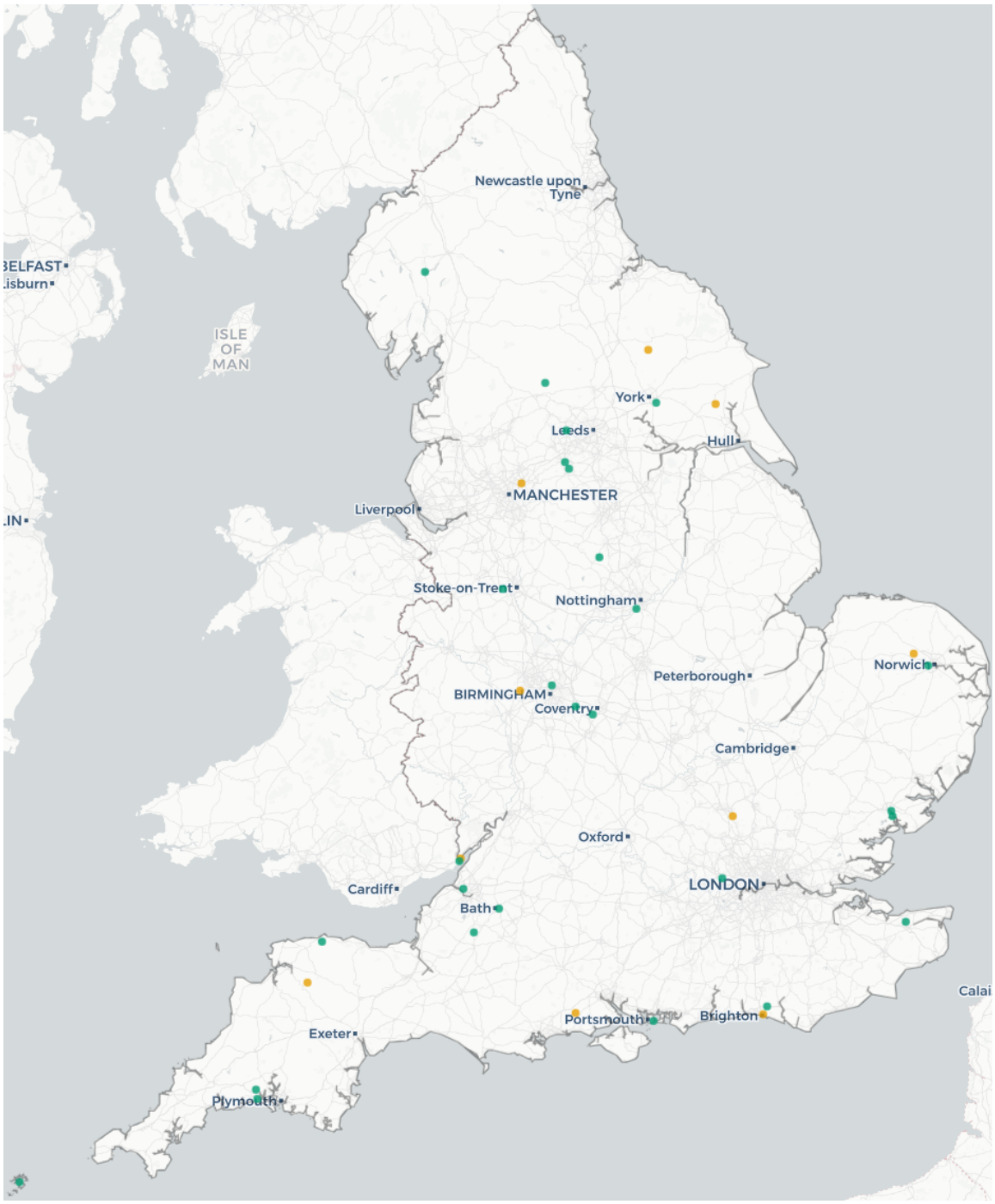


Figure 5.13: The locations of the 37 identified potential singleton outliers on a map of England. The green circles mark local outliers with a low observed prevalence compared to the average observed prevalence of their neighbours, and the orange circles mark local outliers with high observed prevalences.

local behaviour. Hence, one can argue that the area with the 38th largest RDOS value should not be identified as an outlier. Note that many of the areas in the larger outlier set of size 116 also do not look very unusual, but no additional plots are presented here for brevity. Based on the visual analysis, the rest of this chapter will consider only the scenario where the areas associated with the 37 largest RDOS values are identified as outliers. Figure 5.13 shows these 37 potential singleton outliers on a map

of England, where the coloured circles mark the areas' centroids. The green circles mark local outliers with a low observed prevalence (compared to the average observed prevalence of the neighbouring areas over which the RDOS values were computed), and the orange circles mark local outliers with high observed prevalences.

5.5.2 Do the identified outliers appear to be contextual or global?

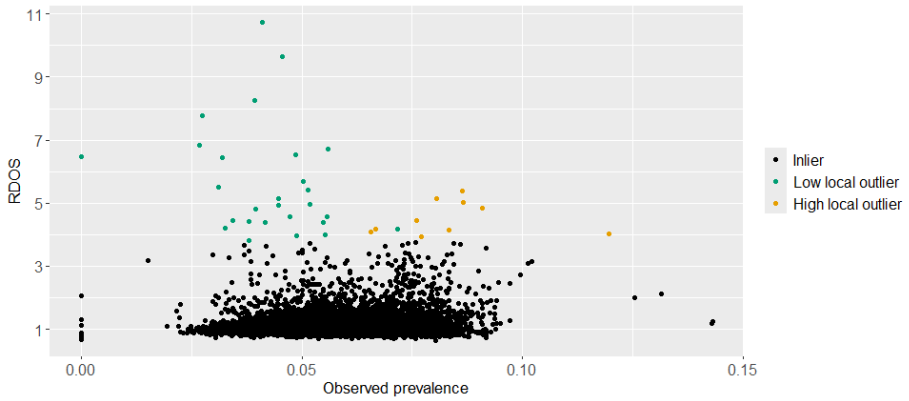


Figure 5.14: A scatterplot showing the RDOS values against the observed asthma prevalence with indications of inlier and outlier sets. The green circles mark local outliers with a low observed prevalence compared to the average observed prevalence of their neighbours, and the orange circles mark local outliers with high observed prevalences.

This section presents an exploratory analysis of the inlier/outlier sets suggested in Section 5.5.1, i.e., where the outlier set consists of the areas with the 37 largest RDOS values. Figure 5.14 shows a scatterplot of the RDOS values against the observed asthma prevalences with an indication of the previously suggested inlier/outlier sets. Here, the coloured circles again indicate low and high local outliers, compared to the average observed prevalences of the areas in the neighbourhoods that were used to compute the RDOS values. The plot allows checking if the identified outliers appear to be global or contextual. The outlier set contains one observation with an observed asthma prevalence of zero, which is the lowest possible value and much smaller than in most areas. Although there are nine other areas with observed asthma prevalences of zero, these nine areas form a spatial cluster which explains why they were not identified as singleton outliers. Hence, the identified singleton outlier with an observed asthma prevalence of zero can be considered a global outlier. On the higher end, the identified outlier with the largest observed asthma prevalence has

a prevalence of 0.1197. Given that there is a substantial gap to the next smaller observed prevalence of 0.1021 and the bulk of the observed asthma prevalences, the area can be considered a global outlier. However, there are five areas with even larger observed asthma prevalences that were not identified as singleton outliers, which should be investigated further. The other 35 areas that were identified as outliers all appear to be contextual outliers, since their observed asthma prevalences do not appear unusual when viewing the dataset as a whole.

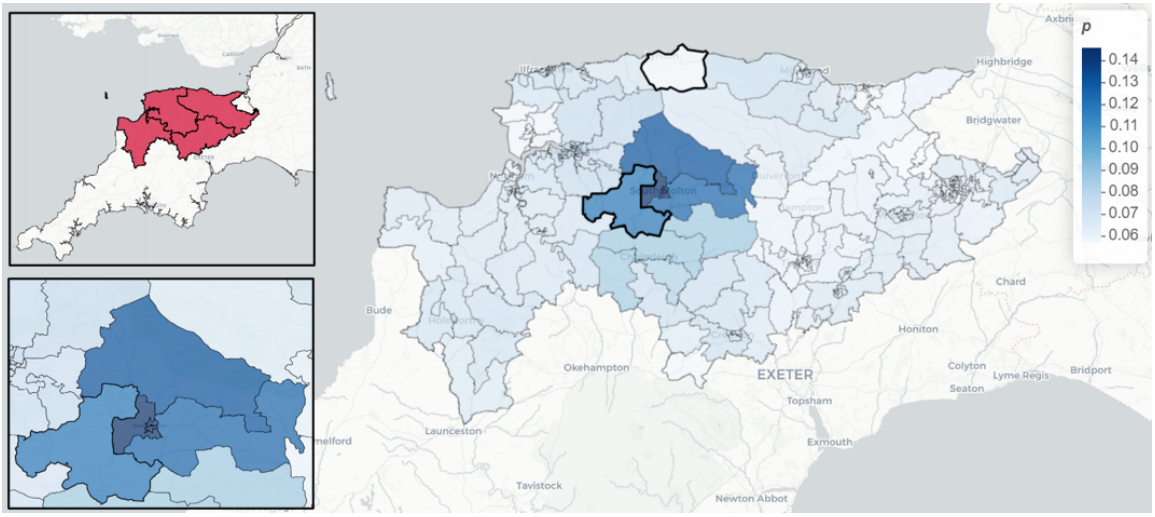


Figure 5.15: A map showing the five LSOAs with the highest observed asthma prevalences.

From Figure 5.14, the question arises why the areas with the five largest observed asthma prevalences were not identified as singleton outliers. Figure 5.15 shows a map with the five areas with the largest observed asthma prevalences, which are located in the local authority district North Devon in the South West of England (the upper boxed image shows the LADs containing the LSOAs presented on a larger part of the South West of England, where the LAD North Devon is the central one of the three Northern LADs). The map reveals that the five LSOAs with the largest observed asthma prevalences form a spatial cluster, as shown up-close in the lower boxed image. The identified outlier is located at the South-Western end of the cluster and has an observed asthma prevalence that is noticeably smaller than those of the areas in the cluster but is substantially larger than in most other areas (recall Figure 5.14), which explains its larger RDOS value and why it was identified as a singleton outlier. Figure 5.14 contains another area that was identified as a singleton outlier, which is located at the Northern shore of the LAD North Devon. The area has an observed

asthma prevalence of 0.0558, which is noticeably smaller than that of its nearby areas (e.g., the areas sharing a border with this area have observed asthma prevalences of 0.0647, 0.0667, 0.0728, and 0.0767) and hence, one can argue that it was correctly identified as a singleton outlier.

5.5.3 How do the potential outliers affect the modelling?

As the final step of the analysis, the modified model BYM2-O is applied to estimate the true asthma prevalence for the entire study region, by accounting for the 37 identified outlying areas. The simulation study in Section 5.4.5 showed that when a larger number of outliers is present, or (at least) a few outliers have a larger outlier magnitude, the modified model can produce overall better point estimates (with a lower median absolute error, see Figure 5.6) and narrower credible intervals (see Figure 5.8) than the conventional BYM2 model, while retaining a coverage close to the theoretical correct value (see Figure 5.7). However, it is important to note that in the simulation study, the study region consisted of 298 areas and the simulation scenarios considered outlier proportions of approximately 1% and 5%. In contrast, in the asthma prevalence application, the study region consists of 32,844 areas and the set of identified outliers is of size 37. Thus, the proportion of identified outliers in the application is approximately 0.11%, which is substantially smaller than in the considered simulation scenarios.

Table 5.3 shows summaries for the fitted BYM2 and BYM2-O models. Specifically, the table shows the posterior mean of the precision parameter τ and mixing parameter ρ of the spatially correlated random effects, with 95% credible intervals shown in parantheses. For the BYM2-O model, the table also shows the posterior mean and 95% credible interval of the precision parameter κ of the independent random effects of the identified outliers. Lastly, the table shows the DIC and WAIC values for the two models.

Table 5.3: Model summaries for the conventional BYM2 and the modified BYM2-O models for the asthma prevalence application.

Model	τ	ρ	κ	DIC	WAIC
BYM2	83.30 (80.35,86.34)	1.00 (1.00,1.00)	-	231,091.30	227,416.18
BYM2-O	86.03 (82.78,89.07)	1.00 (1.00,1.00)	6.49 (3.89,9.97)	230,729.04	226,889.15

An important finding from Table 5.3 is that the mixing parameter ρ has a posterior mean value of 1.00 with a 95% credible interval of (1.00,1.00) in both models, which implies that the BYM2 models of the spatially correlated random effects simplify to ICAR prior distributions (recall Section 2.6.1) that enforce strong spatial autocorrelation. Furthermore, the estimates of the precision parameter τ of the spatially correlated random effects are similar for the two models, with a posterior mean and 95% credible interval (in parantheses) of 83.30 (80.35,86.34) for the BYM2 model and slightly larger values of 86.03 (82.78,89.07) for the BYM2-O model. The larger precision for the BYM2-O model is due to the removal of the potential singleton outliers. Since the potential outliers do not appear to follow their typical local behaviour, they corrupt the estimation of the smooth prevalence surface by inflating the variance (and equivalently, reducing the precision). Since the parameter estimates of the correlated random effects of the two models are fairly similar, the differences between the estimated prevalence surfaces are anticipated to be small. Since the identified outliers are estimated independently from the prevalence surface under the BYM2-O model, the biggest differences are anticipated to be observed for the areas that were identified as outliers and their neighbouring areas. It should be noted that the precision κ of the independent random effects assigned to the identified outliers is estimated to be much smaller than the precision τ of the correlated random effects, with a posterior mean of 6.49 and a 95% credible interval of (3.89,9.97), so the credible intervals of the fitted prevalence of the identified outlying observations are anticipated to be substantially wider than for the identified inlying observations. Since the parameter κ is estimated from a relatively small number of observations, the prevalence estimates of the identified outliers might not be particularly good, and their independent estimation is a sacrifice for obtaining better estimates of the smooth risk surface. The provided DIC and WAIC values in Table 5.3 suggest that the BYM2-O model fits the data slightly better than the BYM2 model, with a DIC value of 230,729.04 compared to 231,091.30 and a WAIC value of 226,889.15 compared to 227,416.18.

The main difference between the conventional BYM2 model and the BYM2-O model is in the estimation of the asthma prevalence for the 37 areas that were identified as singleton outliers and the areas that are neighbours of these areas under

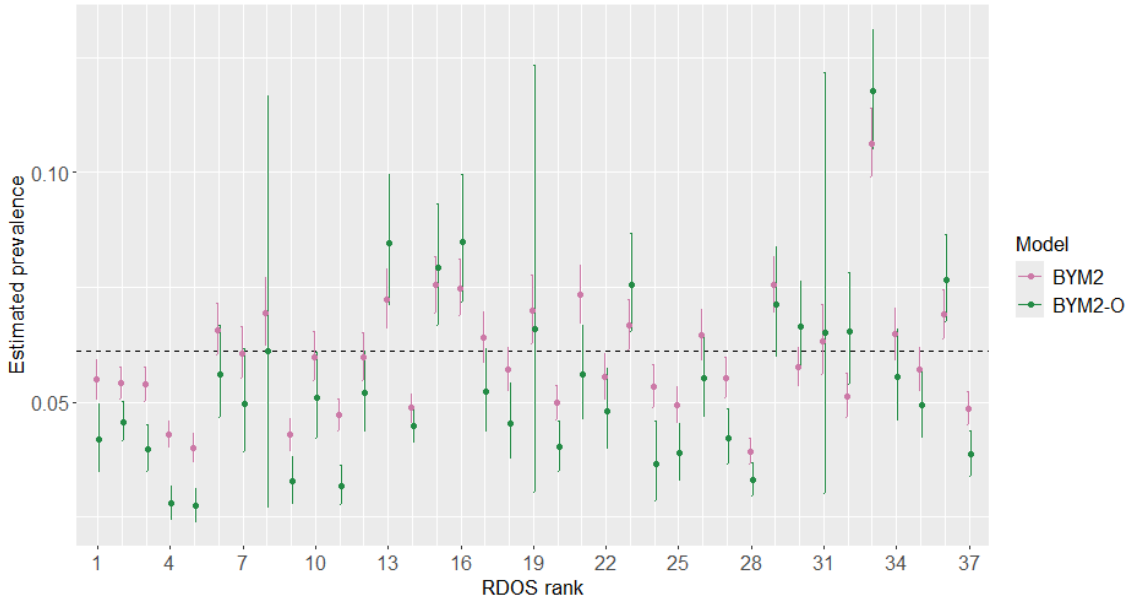
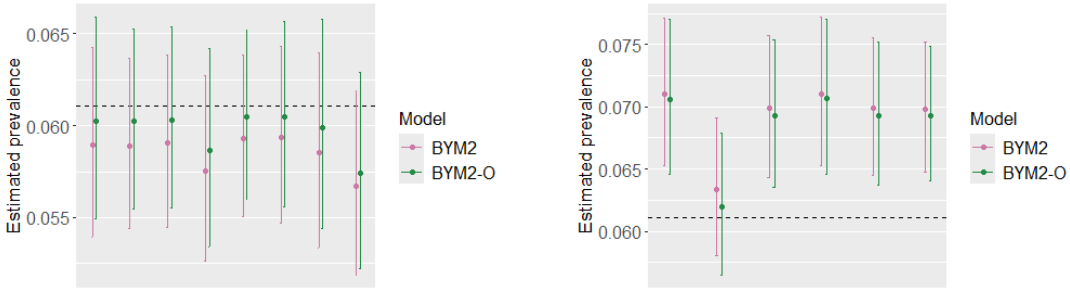


Figure 5.16: The posterior mean fitted asthma prevalence with 95% credible interval for the areas with the 37 largest RDOS values.

the conventional BYM2 model. Figure 5.16 shows that for the conventional BYM2 model, the mean fitted values and 95% credible intervals of the estimated prevalences of the 37 identified outliers tend to be closer to the median observed asthma prevalence (represented by the horizontal dashed line). Under the BYM2-O model, the asthma prevalence in these areas is estimated to be more extreme (i.e., they are impacted less by the smoothing), and for 13 of the 37 areas, the 95% credible intervals obtained from the BYM2-O and BYM2 models do not overlap (i.e., the models provide different estimates for these areas). Another striking observation is that the 95% credible intervals of the asthma prevalence in the areas with the 8th, 19th, and 31st largest RDOS values are much wider under the BYM2-O model than the conventional BYM2 model. The numbers of registered patients in these areas are four, eleven, and twelve with incidence counts of zero, one, and one, respectively. Hence, the wider credible intervals under the BYM2-O model reflect that there is greater uncertainty in the prevalence estimates for these areas with very small populations when the random effects are assumed to be independent of those in the neighbouring areas, while the credible intervals under the BYM2 model are narrower since the assumed strong autocorrelation reduces the variability by borrowing strength from the neighbouring estimated random effects.

The posterior prevalence estimates also differ noticeably for the inlying areas that



(a) The neighbouring areas of the area with the largest RDOS value.

(b) The neighbouring areas of the area with the 13th largest RDOS value.

Figure 5.17: The posterior mean fitted asthma prevalence with 95% credible interval for areas neighbouring outlying areas.

have outlying areas as neighbours, according to the neighbourhood structure applied for the BYM2 model. In this study, an area's neighbourhood consists of its six nearest neighbours (measured as the Euclidean distance between the areas' centroids) and additional neighbours are added so that if area i is in the neighbourhood of area j, then area j is also included in the neighbourhood of area i. Figures 5.17a and 5.17b show the posterior mean fitted asthma prevalences with 95% credible intervals for inlying areas neighbouring the outlying areas with the largest and 13th largest RDOS values, respectively. The area with the largest RDOS value has an unusually low observed asthma prevalence of 0.0409 (compared to the median observed prevalence of 0.0611). The conventional BYM2 model estimates the prevalence in the neighbouring areas by smoothing over the potential outlier, while the BYM2-O model estimates the prevalence in the potentially outlying area independently. As a result, the posterior mean fitted prevalences of the inlying neighbouring areas presented in Figure 5.17a are lower under the BYM2 model than the BYM2-O model. Similarly, the area with the 13th largest RDOS value has an unusually high observed asthma prevalence of 0.0863. As a result, the posterior mean fitted prevalences in the inlying neighbouring areas presented in Figure 5.17b are higher under the BYM2 model than the BYM2-O model.

Figure 5.18 shows boxplots that show the distributions of the absolute differences between the mean fitted prevalences from the BYM2 and BYM2-O models over inliers not neighbouring outliers, inliers neighbouring outliers, and outliers. The largest differences in the estimated prevalences between the BYM2 and BYM2-O models can be observed for the identified outliers (with a median absolute difference of 0.0096),

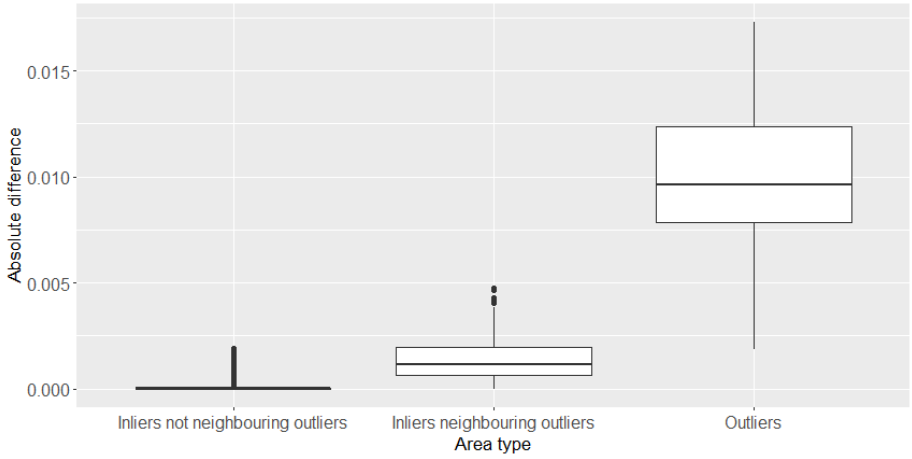


Figure 5.18: Boxplots showing the distributions of the absolute differences between the mean fitted prevalences from the BYM2 and BYM2-O models over inliers not neighbouring outliers, inliers neighbouring outliers, and outliers.

which are pulled closer to their neighbouring estimates under the BYM2 model and estimated to be more extreme under the BYM2-O model. Similarly, noticeable differences can be observed for the prevalence estimates of the inlying areas that are neighbours of the identified outliers (with a median absolute difference of 0.0012). In contrast, the two models' prevalence estimates of the inlying areas that are not neighbours of the identified outliers tend to be almost identical (with a median absolute difference less than 0.0001).

5.6 Discussion

The study presented here provides a generalisation of the relative density-based outlier score (RDOS) of Tang and He (2017) to be applicable to areal data with non-spatial attributes. The proposed RDOS values are used for singleton outlier detection for spatially smooth disease data, where the outlier detection's effectiveness is evaluated in a simulation study and shown to provide substantially better detection results than what can be achieved using the commonly applied local Moran's I statistic. The second key contribution is the proposed novel modified smoothing model BYM2-O, which estimates the disease prevalences in the areas that were identified as outliers independently from the inferred smooth prevalence surface over the inlying areas, which is estimated using a BYM2 model. The proposed method is a two-stage modelling approach, where the potential outliers are identified from the RDOS values in

the first stage to provide the inlier and outlier sets used by the modified smoothing model in the second stage. The method's effectiveness is again evaluated in a simulation study, for which the modified smoothing model BYM2-O produces better estimates than the conventional BYM2 model when the number of outliers is large or (at least) a few outliers with larger outlier magnitudes are present in the data.

The relative density-based outlier score (RDOS) proposed in this study is computed from local kernel densities. These kernel densities require the specification of a bandwidth parameter. The studies by Schubert et al. (2014a) and Tang and He (2017) also use kernel-based methods but ignore their methods' sensitivity with regard to the bandwidth parameter specification. The study presented here fills that gap by proposing an algorithm for specifying the bandwidth parameter value, where the bandwidth is increased incrementally until the ranking of the computed RDOS values stabilises. Here, the similarity between rankings is computed using Kendall's rank coefficient and a threshold of 0.99 is chosen such that the smallest bandwidth is selected for which two subsequent rankings have a Kendall's rank coefficient greater than 0.99. The simulation studies have shown that the RDOS values computed for the so obtained bandwidth parameter provide excellent detection results, as shown via inspection of the receiver operator characteristic (ROC) curves.

The simulation study showed that the RDOS values tend to be highest for the true singleton outliers, especially when the outlier magnitude is large. However, a limitation of the method is the uncertainty around choosing the number of observations to be identified as outliers. The partitioning around medoids (PAM) algorithm does not generally provide good splits into inlier and outlier sets, as the automated method tends to provide conservative results (only a subset of the true outliers are identified) when many outliers of large outlier magnitude are present and the method is highly sensitive to the number of RDOS values considered by the algorithm when the proportion of outliers or the outlier magnitude are small. Hence, one should investigate the observations with the largest RDOS values to check which of them appear to be potential outliers, as demonstrated in the asthma prevalence application presented in this study. The limitation of the manual approach is that it is subjective, as it is up to the researcher to decide which observations should be considered potential

outliers. At the same time, the method's subjectivity gives greater flexibility to the user, which means that expert knowledge can be considered when investigating the potential outliers.

For the simulation scenarios considered in this study, the proposed modified smoothing model BYM2-O was shown to provide overall better estimates of the true prevalence than the conventional BYM2 model when a larger number of outliers or some outliers of large magnitude are present in the data, even when the automated outlier detection that uses the PAM algorithm to split the RDOS values into inliers and outliers results in sub-optimal outlier sets. However, when only few small outliers are present in the data, the conventional BYM2 model tends to provide better estimates and hence, the modified model should only be applied when the observations in the identified outlier sets appear to differ substantially more from their neighbouring values than the typical variability observed in most of the data.

The motivating application of asthma prevalence data at a small-area LSOA level in England showed that individual areas that deviate from the typical local behaviour can be identified using the singleton outlier detection method using the RDOS values. In contrast, the method is not designed to identify collective outliers, as demonstrated by the collective outliers that could be identified from further inspection in this application. Hence, other methods should be consulted when the objective is to identify collective outliers. The posterior mean fitted values of the asthma prevalence in some of the 37 identified outlying areas differ strongly for the BYM2 and BYM2-O models, with the BYM2 model providing estimates that tend to be closer to the overall observed median asthma prevalence. In comparison, the posterior mean fitted values obtained from the BYM2-O model tend to be further away from the overall observed median asthma prevalence, since the values are estimated without assuming that the prevalence in these areas should be similar to their neighbouring estimates. The differences between the posterior mean prevalences of the inlying areas tend to be marginal, but noticeably larger differences can be observed for the inlying areas that have an identified outlier in their neighbourhood. The marginal differences between most inlying areas are likely due to the dataset's very large size; i.e., estimating the parameters that determine the smooth risk surface over 32,807 instead of the complete

32,844 areas leads to fairly similar results. Bigger impacts might be observed for smaller datasets, such as the ones considered in the simulation studies.

Some data limitations should be identified for the motivating application; recall that the asthma prevalence is computed from the number of patients who were prescribed asthma-related drugs in the preceding 12 months, although GP practices reporting asthma prevalence beyond 30% are excluded from the calculation. Firstly, the asthma prevalence captures only registered patients and hence, asthma cases in individuals that do not seek medical attention are not captured. As a result, factors impacting the proportion of individuals that are registered and seeking medical attention will likely have an influence on the observed asthma prevalence. Secondly, it is unclear what impact the exclusion of GP practices reporting asthma prevalence beyond 30% might have on the analysis. If these were genuine asthma cases, the reported asthma prevalences in the corresponding areas might be too small. Lastly, the number of registered patients are quite small in some areas, which could lead to unusual observed asthma prevalences; e.g., recall the ten areas with zero asthma patients - in these areas, the numbers of registered patients range from two to eight people. However, Figure 5.14 showed that 35 of the 37 identified outliers appeared to be contextual and that only one of the identified outliers that appeared to be a global outlier had an observed asthma prevalence of zero, while the other nine areas with an observed prevalence of zero form a spatial cluster and were not identified as singleton outliers. Hence, the small numbers of registered patients in some areas do not appear to have been the main reason behind the unusual observed asthma prevalences in most of the identified singleton spatial outliers.

Due to the novelty of the presented methods, there is much room for future work. While the RDOS values can be computed for any areal observations whose spatial structure can be described by a neighbourhood matrix, this study only evaluated the method's efficacy for disease prevalence data generated from a binomial logit-linear model. Here, the generated data depend only on the population sizes and the disease prevalence surface. Another parameter that is often the target of inference in disease mapping applications is disease risk, which is computed from the observed and expected counts. Hence, one could reproduce the presented study for outlier detection

in standardised morbidity or mortality rates, which depend on the expected counts and the disease risk surface, and could be generated from a Poisson log-linear model. Similarly, the modified smoothing model BYM2-O can be evaluated for a likelihood other than the binomial distribution.

Another natural extension to both the RDOS values and the BYM2-O model is to make these methods spatio-temporal. In this study, the methods were proposed for the spatial domain, but one might be interested in how the values of the potential outliers change over time or identify areas with unusual temporal trends. Hence, the proposed methods could be extended to be able to identify spatio-temporal outliers and to account for the potential outliers in the modelling of spatio-temporal data.

Lastly, since the PAM algorithm does not appear to be a reliable method for identifying the inlier and outlier sets automatically, a possibility for future work is to investigate or develop other automated methods for splitting the RDOS values into the two sets. These methods should be capable of identifying smaller outlier sets even when all observations are considered and the RDOS values of some outliers are much larger than those of the rest of the data. Additionally, the PAM algorithm applied in this study identifies at least one observation as an outlier. Hence, another possibility for future work is to propose a method that can result in the conclusion that no observations should be identified as outliers.

Chapter 6

Discussion and future work

The work presented in this thesis contributes to the statistical literature in spatio-temporal epidemiology and disease mapping. Chapters 3 and 4 provide timely analyses of important public health aspects of the COVID-19 pandemic in the UK, and Chapter 5 proposes novel statistical methods to identify and account for potential singleton spatial outliers in areal data applications. This chapter summarises the three studies' main aims and key findings, followed by a discussion of limitations and ideas for future work.

The study presented in Chapter 3 investigates the spatio-temporal patterns and trends of COVID-19 mortality risk in England to identify geographical differences in the impact of national lockdown. It answers four main questions:

1. How long after the implementation of lockdown did mortality risks reduce at a national level, and did this vary by lockdown?

2. How did the temporal trends in mortality risks differ by region in England?

3. Which local authorities were exposed to the highest average risks in the weeks after lockdown?

4. Which local authorities shared similar temporal trends in mortality risks?

The key findings are that the mortality risks increased drastically before the implementation of lockdowns 1 and 3; they stopped increasing three weeks into lockdown

and reduced to pre-lockdown levels after ten and six weeks, respectively (see Figure 3.5). There was no drastic increase in mortality risk before the implementation of lockdown 2, and the second lockdown saw no meaningful reduction in mortality risk (possibly because it was too short, at only four weeks). It should be noted that England had implemented a three-tier system of mobility restrictions that started on 14th October before lockdown 2, with varying degrees of restrictions for the LADs in the three tiers, which Davies et al. (2021) suggested has had a sizeable effect in reducing the number of COVID-19 deaths. Hence, the preceding tiered restrictions might explain the reduced impact of lockdown 2. The analysis of temporal risk trends at a regional level showed that although some regional differences in risk trends are apparent, there was no clear hierarchy for the nine regions of England across the three lockdowns (see Figure 3.6). Identifying areas with the highest average risks in the weeks after lockdown suggested that the local authorities at the highest risk appear to be deprived districts within regions at higher risk (see the discussion of Table 3.1). Lastly, identifying local authorities with similar temporal trends revealed that in lockdown 1, there appeared to be an urban/rural divide in mortality risks, with the urban areas tending to have a higher peak risk (see Figure 3.9a). In lockdown 3, the peak risk appeared higher in southeast England (see Figure 3.9b), associated with the early spreading of the Alpha variant in that region.

Chapter 4 presents an epidemiological analysis of the COVID-19 vaccine uptake over the first three doses by sex and age group across Scotland's council areas, answering three main questions:

1. Are there any trends in attrition rates by age group, and does this vary by sex?

2. Are there any spatial patterns in attrition rates by council area, and does this vary by sex?

3. How do these trends and patterns compare across the two transitions (from doses 1 to 2 and 2 to 3)?

The first key finding is a strong association between age and attrition rates (see Figure 4.3), where the odds in favour of attrition decrease smoothly with increasing age (i.e., older age groups are more likely to continue receiving vaccine doses). Overall,

males tend to have approximately 1.1 times higher odds of attrition than females. The variation in the odds of attrition between age groups tends to be greater for males, and the decreasing effect of older age on the odds of attrition appears smaller for females (see Figure 4.3). Secondly, while the attrition rates for the two sexes can differ substantially in some council areas, the council areas with the highest and lowest attrition rates tend to have attrition rates that are consistently higher/lower for both sexes (see Figure 4.5). Further investigation of these council areas suggests that there might be a positive association between deprivation and attrition rates, meaning that more deprived areas tend to have higher attrition rates and vice versa. Regarding the differences between transitions, the study shows that the variation in the odds of attrition between age groups tends to be overall greater in the second transition than in the first; otherwise, the smoothly decreasing odds by increasing age group can be seen for both transitions (see Figure 4.3). The average odds of attrition over all age groups and council areas in Scotland are 4.1 times higher in the transition from doses 2 to 3 than from doses 1 to 2, which suggests increased vaccine fatigue over time. Lastly, the council areas with the highest and lowest attrition rates tend to have attrition rates that are consistently higher/lower for both transitions (see Figure 4.4).

A common key finding from Chapters 3 and 4 is that the spatial smoothness assumption might not hold consistently across the whole study region (e.g., Chapter 3 showed an urban/rural divide in peak risk during the first lockdown, and Chapter 4 identified an unusual observation as a potential outlier), which will impact the estimation whenever a global smoothing model is applied. Motivated by the finding that the observations in some areas might violate the spatial smoothness assumption, the study in Chapter 5 proposes novel methodology for identifying singleton spatial outliers in large datasets (e.g., with thousands of areal units) and proposes a modified spatial smoothing model BYM2-O that accounts for these identified potential outliers. The efficiency of the proposed relative density-based outlier score (RDOS) and the modified smoothing model are evaluated in simulation studies. Firstly, these simulations show that the RDOS values can be used to identify singleton outliers effectively, with much better detection results than the commonly used local Moran's

I statistic (see Figure 5.4). Secondly, the modified smoothing model BYM2-O produces overall better estimates than the conventional BYM2 model when the number of outliers is large or at least some of the outliers have a large magnitude (see Figures 5.5 to 5.8), even when the identified outlier set is not exactly the same as the set of the induced outliers in the simulated data.

The outlier detection method and modified smoothing model are applied with a two-stage modelling approach in a motivating study of asthma prevalence at England's lower super output area (LSOA) level. The motivating study aims to answer three main questions:

1. Which areas appear to be potential singleton outliers?

2. Do the identified outliers appear to be contextual or global?

3. How do the potential outliers affect the modelling?

The first key finding is that 37 areas can be identified as potential outliers, and there are no obvious patterns regarding the locations of these areas on the map of England (see Figure 5.13). Secondly, two of the 37 areas appear to be global outliers (one has an observed prevalence of zero, and the other is among the areas with the largest observed prevalences); the other 35 areas appear to be contextual outliers (i.e., their observed prevalences deviate from their typical local behaviour but do not appear unusual when viewing the whole dataset) - see Figure 5.14.

Lastly, to identify how the potential outliers affect the modelling, the data are fitted using two hierarchical binomial logit-linear models where the spatially correlated random effects are estimated via the modified smoothing model BYM2-O (which estimates the identified potential outliers independently from the smooth prevalence surface) and the conventional BYM2 model (which estimates the smooth prevalence surface over all areas), respectively. The prevalence estimates from these two models are similar for the inlying areas that do not have outlying areas in their neighbourhoods (see Figure 5.18). In contrast, there are substantial differences in the prevalence estimates for the outlying areas (see Figure 5.16), and the estimates for the inlying areas that have outlying areas in their neighbourhoods also differ noticeably

for the two models (see Figure 5.17). The differences in the prevalence estimates highlight the importance of considering potential outliers in the analysis of areal data. Furthermore, if at least some of the potential outliers were correctly identified, the simulation studies suggest that the prevalence estimates obtained from the BYM2-O model should be better than those from the BYM2 model, overall.

There are some limitations common to the three studies presented in this thesis. Firstly, the presented studies are all observational. Hence, while these studies identified associations, trends, and patterns, they did not detect any 'causal effects' (e.g., the possible impact of lockdown on mortality risk is associative, since the counterfactual event of what would have happened without the lockdown could not be observed). Although some considerations in attempting to distinguish causal effects from possibly non-causal associations in epidemiology were proposed and further considered (e.g., see Hill, 1965 and Rothman et al., 2008), they cannot guarantee that the observed associations are not due to some other factors not considered in the analysis. For example, Hill (1965) states that "none of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis". Since active research continues to seek to identify causal relationships in epidemiological studies (e.g., see Vandenbroucke et al., 2016 or Cox, 2021), it should be clarified that the work presented in this thesis made no restrictive assumptions to attempt inferring causality and the identified relationships are all associative.

Another point of clarification is that the presented analyses are applications of descriptive epidemiology - they estimate disease risk and prevalence but do not aim to measure associations between health outcomes and ecological (e.g., temperature, air pollution, ethnicity) or socio-economic (e.g., deprivation, education level, unemployment) covariates. Still, possible associations between the estimated risk or prevalence and other factors might be identified retrospectively (e.g., recall the identification of a possible positive association between deprivation and higher peak risks during lockdown 1 in the study presented in Chapter 3), or be investigated further based on the presented results (e.g., one could identify potential risk factors by inspecting the identified outliers in Chapter 5).

Lastly, note that all three studies analysed data aggregated at the area level.

Using aggregated data implies some limitations. For example, people moving home between areas might result in data errors such as duplicate or missing observations between subsequent time points (e.g., a person might have received a first dose of the COVID-19 vaccine while registered in one area and moved home to a different area before receiving the second dose). Also, the results depend on the borders that determine which individuals are grouped, which is known as the *modifiable areal unit problem* (Openshaw, 1983). The key idea is that, ideally, disease risk or prevalence should be represented as a continuous surface so that each point on the map has a specific value. In contrast, disease data are often aggregated at an area level, and the disease risk or prevalence is estimated to be constant in each area, with jumps in the estimates between areas. The areal units at which the data are aggregated are typically designed for multiple uses and not necessarily optimised for aggregating health data. For example, the local authority districts in England (the areal units in Chapter 3) are used for local government (e.g., see Ministry of Housing, Communities and Local Government, 2023). Since the data were only provided for the specific areal units, it is important to note the limitations associated with the modifiable areal unit problem, and for further discussion, see, for example, Manley (2021).

The studies presented in this thesis provide many opportunities for future work. An obvious extension is to pose the motivating questions that were answered for applications in Scotland and England for data from other countries. For example, the study presented in Chapter 3 investigated which local authorities were exposed to the highest average risks in the weeks after lockdown. A key finding was an urban/rural divide in mortality risks in lockdown 1, with the urban areas tending to have a higher peak risk. Bayerlein (2024) referenced the presented study and highlighted this finding as a similarity to data from the European Union, where the peak in excess mortality was most marked in urban areas during the first wave of the pandemic. Similarly, the other motivating questions could be investigated for comparison with other countries, possibly even in a meta-analysis combining the findings from existing studies to estimate a worldwide impact of lockdowns.

The key findings presented in this thesis might also inform future public health concerns in the UK. For example, Naylor et al. (2024) referenced the study presented

in Chapter 4 to hypothesise that in the UK vaccination programme against the respiratory syncytial virus, clinicians may encounter vaccine hesitancy, especially among men and those from more deprived backgrounds, while older people could be more likely to accept the offered vaccines. Though the identified associations presented in this thesis for the COVID-19 vaccines might not be the same for other vaccines or consistent over time, the awareness of possible associations that can be considered is valuable for such related applications.

There are also many opportunities for future work on methodological extensions to the studies presented in this thesis. For example, one could extend the spatio-temporal analysis presented in Chapter 3 to a multivariate setting and simultaneously analyse mortality risks and hospitalisation rates after national lockdowns. Similarly, one could extend the study presented in Chapter 4 by simultaneously modelling attrition rates for different types of vaccines with multiple doses (e.g., the MMR, varicella (chicken pox), or influenza vaccines) to evaluate if there are differences in the identified vaccine fatigue trends. Especially for the study presented in Chapter 5, there are many opportunities for future work on methodological extensions. The identification of potential singleton spatial outliers and the modified spatial smoothing model motivate further methodological research questions, such as:

- How can temporal outlying areas be identified? (e.g., how might one identify an area whose observed prevalence is unusual at a particular point in time?)
- How can spatio-temporal outlying areas be identified? (e.g., how might one identify an area whose observed prevalence is unusual at a particular point in time, compared to the observed prevalence of its neighbouring areas at that time point?)
- How can areas with unusual temporal trends be identified? (e.g., how might one identify an area whose change in observed prevalence over time differs substantially from that in its neighbouring areas?)
- How could a spatio-temporal model be modified to account for such unusual temporal trends?

Appendix A

Additional Analysis for Chapter 3

This appendix provides convergence and posterior predictive checks, and additional analyses for Chapter 3. Section A.1 provides traceplots of the simulations from the posterior densities of the MCMC algorithm, Section A.2 provides posterior predictive checks, and Section A.3 provides a sensitivity analysis on the prior choice of the variance parameter of the spatio-temporal random effects presented in Equation (3.13) of Chapter 3.

A.1 Graphical convergence checks for the MCMC algorithm

Figure A.1 presents trace plots for the AR(2) model to illustrate the convergence, and they do not show strong evidence against convergence (the plots for the AR(1) model look similar). While the parameter spaces for ρ, α_1, α_2, and τ^2 were explored quickly, the trace plots for β_0 and ϕ_{11} suggest that the algorithm was slower to explore the parameter spaces for these parameters. However, especially for the intercept term β_0 this is not an uncommon observation and hence should not cause any greater concern. Note that there are 20,904 (312×67) spatio-temporal random effects in the model; hence Figure A.1 only contains the trace plot for one of these random effects, ϕ_{11}, for brevity.

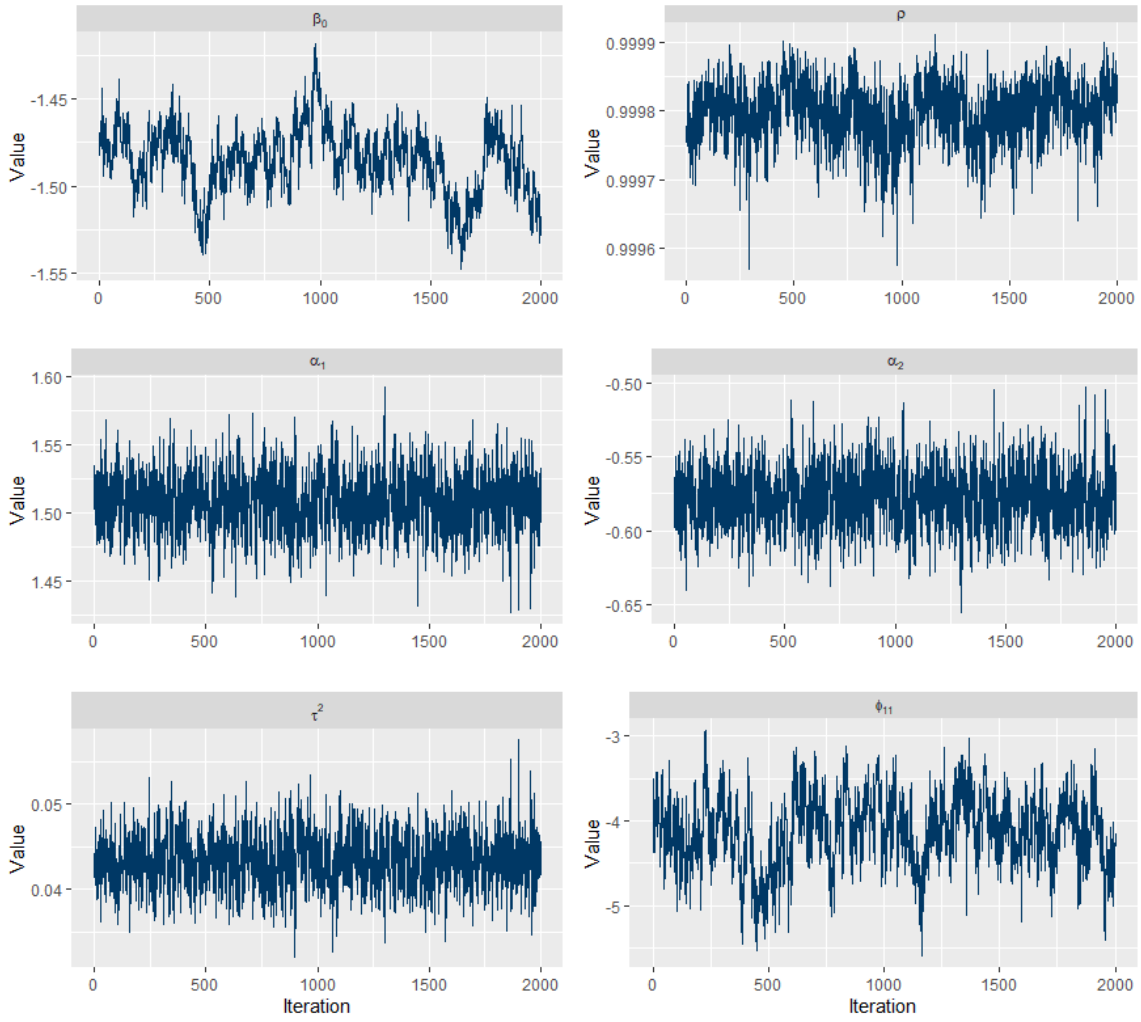


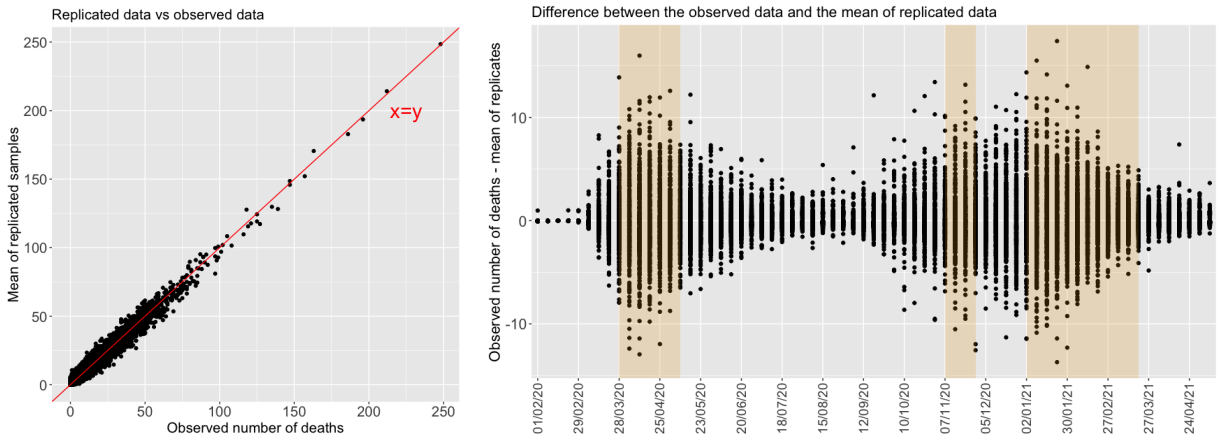
Figure A.1: Traceplots for the simulated values of β_0, ρ, α_1, α_2, τ^2, and ϕ_{11} in the AR(2) model.

A.2 Posterior predictive checks

The MCMC algorithm produced 2,000 posterior samples for each parameter, from which replicated samples of the observed number of deaths were generated from the Poisson distribution that was assumed in model (3.7). That is, for each observation Y_{kt}, replicated observations $Y_{kt}^{rep(i)} \sim \text{Poisson}(e_k\tilde{\theta}_{kt}^{(i)})$ were generated for $i = 1, \ldots, 2000$.

The points in Figure A.2a show the observed Y_{kt} against the mean of the corresponding replicated observations $\overline{Y}_{kt}^{rep} = \frac{1}{2000}\sum_{i=1}^{2000} Y_{kt}^{rep(i)}$, for $k = 1, \ldots, 312$ and $t = 1, \ldots, 67$. The weeks coloured in beige comprise the lockdowns. The plot indicates that the means of the replicated observations are relatively similar to the observed number of deaths, which suggests that the model fits the data reasonably well. Figure A.2b shows the difference between the observed number of deaths and

the mean of the corresponding replicated observations, i.e. $Y_{kt} - \overline{Y}_{kt}^{rep}$. These differences are plotted by week and it looks as though the differences are larger in the weeks when the observed number of deaths were higher (as confirmed by Figure 3.3). The larger differences during these weeks are expected, as the observed data contain more variation. Overall, the replicated observations do not seem to be generally overestimating or underestimating the observed number of deaths, which further suggests that the model appears to be appropriate.



(a) Observed number of deaths vs the mean of the replicated observations.

(b) Difference between the observed number of deaths and the mean of the replicated observations, by week.

Figure A.2: Posterior predictive checks: Observed vs replicated observations.

As an alternative posterior predictive check, the following p-values can be computed

- $p_1 = P((Y < T_{0.025}(\boldsymbol{Y}^{rep})) \cup (Y > T_{0.975}(\boldsymbol{Y}^{rep})))$,
- $p_2 = P((Y \leq T_{0.025}(\boldsymbol{Y}^{rep})) \cup (Y \geq T_{0.975}(\boldsymbol{Y}^{rep})))$,

where p_1 is defined with a strict inequality while p_2 uses a non-strict inequality. Here $T_{0.025}(\boldsymbol{Y}^{rep})$ and $T_{0.975}(\boldsymbol{Y}^{rep})$ denote the 2.5th and 97.5th quantiles of the replicated observations, respectively. An overall p-value over all data points of $\tilde{p}_1 = \frac{1}{312\times67}\sum_{k=1}^{312}\sum_{t=1}^{67} I((Y_{kt} < T_{0.025}(\boldsymbol{Y}_{kt}^{rep})) \cup (Y_{kt} > T_{0.975}(\boldsymbol{Y}_{kt}^{rep})))$ is estimated, where $I()$ is an indicator function that takes on the value of 1 if the argument inside the function is true and 0, otherwise. Similarly, the overall p-value $\tilde{p}_2 = \frac{1}{312\times67}\sum_{k=1}^{312}\sum_{t=1}^{67} I((Y_{kt} \leq T_{0.025}(\boldsymbol{Y}_{kt}^{rep})) \cup (Y_{kt} \geq T_{0.975}(\boldsymbol{Y}_{kt}^{rep})))$ is estimated. These estimated p-values take on values of $\tilde{p}_1 = 0.0057$ and $\tilde{p}_2 = 0.1152$. The expected value of 0.05 falls between the two estimates, which is another indication that the model fits the data reasonably well.

A.3 Sensitivity analysis on the prior choice of the variance parameter

The log-transformed mortality risk θ_{kt} is fitted by an overall mean β_0 and a spatio-temporal random effect ϕ_{kt} (see Equation 3.8). In the AR(2) version of the model, the spatio-temporal random effects are assigned a joint prior distribution defined by the parameters τ^2, ρ, α_1, and α_2 (see Equation 3.10). The paramters ρ, α_1, and α_2 are assigned flat prior distributions, and the mean of the log-transformed mortality risk is assigned a weakly informative normal prior distribution, i.e. $\beta_0 \sim N(\mu_0 = 0, \sigma_0^2 = 10,000)$. Thus, when estimating these parameters the model lets the data speak for themselves. The variance parameter τ^2 is assigned an Inverse-Gamma$(a = 1, b = 0.01)$ prior distribution, but other specifications of the shape parameter a and scale parameter b have been proposed previously. Thus a sensitivity analysis is undertaken on this prior specification by refitting the model with $(a = 0.001, b = 0.001)$, $(a = 1, b = 1)$, and $(a = 1, b = 0.1)$. The resulting estimates of τ^2 are presented in Table A.1 below.

Table A.1: Median and 95% credible interval for the simulated values of τ^2 under Inverse-Gamma prior distributions with different parameter values.

a	b	Median	95% CrI	
1	0.01	0.0432	(0.0372,0.0496)	(original model)
0.001	0.001	0.0433	(0.0373,0.0502)	
1	0.1	0.0439	(0.0378,0.0503)	(alternative models)
1	1	0.0484	(0.0420,0.0556)	

The 95% credible intervals of all three alternative models contain the median posterior estimated variance of the original model. Further, the correlation coefficients between the fitted risks from the original model and those of the alternative models all take on a value of 1.0 (to 3 decimal places). The mean absolute differences between the fitted risks of the original model and the alternative ones are computed to assess their similarity, and these take on values of 0.004 (for $a = 0.001, b = 0.001$), 0.005 (for $a = 1, b = 0.1$), and 0.004 (for $a = 1, b = 1$), suggesting consistent risk estimation.

Appendix B

Additional Analysis for Chapter 4

This appendix provides additional plots, derivations, and analyses for Chapter 4. Section B.1 provides additional plots of the observed attrition rates and Section B.2 shows the full map of Scotland, as Figure 4.2 showed a modified map for compactness. Section B.3 shows a derivation of the expectation of attrition rates, given the number of people who received a preceding dose of the vaccine. Derivations for the interpretation of the model's fixed and random effects are presented in Section B.4. Section B.5 shows a plot of observed versus fitted attrition rates to confirm that the model fits the data appropriately, and Section B.6 is a proof of principle simulation which demonstrates that the proposed model can accurately estimate the model parameters from data similar to that of the study in Chapter 4. Section B.7 presents a sensitivity analysis for the hyperprior specifications of the variance and mixing parameters of the random effects and Section B.8 presents a sensitivity analysis for the choice of structure assigned to the random effects' BYM2 priors. Section B.9 shows tables of the posterior means and credible intervals of the odds ratios of attrition for all age groups and council areas. Lastly, Section B.10 provides an additional analysis that considers additional variables to examine their relationships with attrition rates.

B.1 Additional plots for the observed attrition rates

In Section 4.2.4.1, I state that for age groups 75 years and older, males tend to have lower attrition rates than females, which is not immediately clear from Figure 4.1 as it uses one scale for all age groups. So, this section provides additional plots.

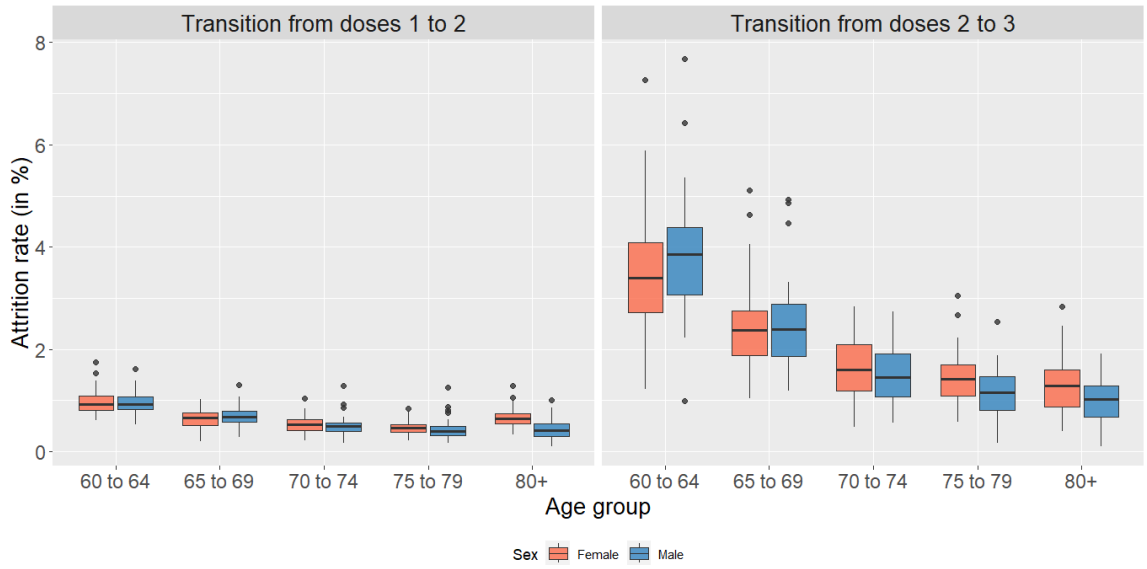


Figure B.1: Attrition rates (in %) for the age groups 60-64, 65-69, 70-74, 75-79 and 80+, by sex and transition (from doses 1 to 2 and 2 to 3).

(a) For the age groups 18-29, 30-39, 40-49, 50-54, 55-59, 60-64, 65-69, and 70-74.

(b) For the age groups 75-79 and 80+ for the two transitions, by sex.

Figure B.2: Scatterplots of the attrition rates (in %).

Figure 4.1 of Chapter 4 shows that attrition rates are higher for males than females in all age groups 59 years and younger. Figure B.1 shows the same boxplots

as Figure 4.1, but only for the age groups 60-64, 65-69, 75-79, and 80+. The figure shows that the attrition rates are similar for the two sexes in age groups 60-64, 65-69, and 70-74 and that the attrition rates tend to be higher for females than for males in the age groups 75-79 and 80+. This reversing trend becomes even clearer when looking at the council area, age group, and transition-specific comparisons of attrition rates between the two sexes presented in Figure B.2. The colour of the points indicate whether an observation is from the first or the second transition. When a point falls on the dashed equality line, the attrition rates are the same for females and males in the corresponding age group, council area, and transition. For a point above the dashed line, the attrition rate is higher for males and if it falls below the dashed line, it is higher for females. Figure B.2a shows that for the age groups 74 years and younger, in both transitions, the attrition rates tend to be higher for males than for females in almost all council areas. Figure B.2b shows that for the age groups 75-79 and 80+, the attrition rates tend to be higher for females than for males (as most points fall below the dashed equality line). Hence, these plots provide a clearer demonstration of the finding in Section 4.2.4.1 that, while males tend to have higher attrition rates than females in the younger age groups, this trend reverses for age groups 75 years and older, where males tend to have lower attrition rates than females.

B.2 Full map of Scotland

Figure B.3: The map of Scotland at full scale.

Figure 4.2 in Section 4.2.4.2 shows maps of Scotland where the Shetland Islands and Orkney Islands are downscaled and presented in a box rather than at their true location in the North of Scotland to make the maps more compact. Figure B.3 is included for reference, showing the map of Scotland at its full scale.

B.3 Derivation of the expectation of attrition rates

This section provides a derivation of the result $\mathbf{E}\Big[A_{sgk}^{(d-1)d} \mid Y_{sgk(d-1)}\Big] = \alpha_{sgk}^{(d-1)d}$, which is stated in Section 4.3.1.

The observed attrition rate is defined as $A_{sgk}^{(d-1)d} = 1 - \Big(Y_{sgkd} \Big/ Y_{sgk(d-1)}\Big)$, for $d = 2, 3$, where Y_{sgkd} is the number of people of sex s, age group g, and council area k who receive dose d of the vaccine. I assume that Y_{sgkd} can be modelled with the binomial likelihood, as

$$Y_{sgkd} \mid Y_{sgk(d-1)} \sim \text{Binomial}\Big(n = Y_{sgk(d-1)}, p = 1 - \alpha_{sgk}^{(d-1)d}\Big). \tag{B.1}$$

Hence, the expectation of the attrition rate, given the number of people who received the preceding dose, can be computed as

$$\begin{aligned}
\mathbf{E}\Big[A_{sgk}^{(d-1)d} \mid Y_{sgk(d-1)}\Big] &= \mathbf{E}\left[1 - \frac{Y_{sgkd}}{Y_{sgk(d-1)}} \mid Y_{sgk(d-1)}\right] \\
&= 1 - \frac{\mathbf{E}\big[Y_{sgkd} \mid Y_{sgk(d-1)}\big]}{Y_{sgk(d-1)}} \\
&\overset{\text{(B.1)}}{=} 1 - \frac{Y_{sgk(d-1)} \times \Big(1 - \alpha_{sgk}^{(d-1)d}\Big)}{Y_{sgk(d-1)}} \\
&= \alpha_{sgk}^{(d-1)d}.
\end{aligned}$$

B.4 Derivations for the interpretation of the fixed and random effects

This section provides derivations for the interpretation of the fixed and random effects of Model 1, but the effects of Model 2 can be interpreted similarly. The odds in favour of attrition for Model 1, with transition-specific council area effects that are the same for both sexes is given by

$$\alpha_{sgk}^{(d-1)d} \Big/ \left(1 - \alpha_{sgk}^{(d-1)d}\right) = \exp(-\beta_0) \times \exp(-\gamma_s) \times \exp(-\psi_d) \times \exp\left(-\delta_g^{(sd)}\right) \times \exp\left(-\phi_k^{(d)}\right).$$

The term $\exp(-\beta_0)$ is the geometric mean (GM) of the odds in favour of attrition for females across all age groups and council areas in the first transition (i.e., for $\gamma_s = 0$ and $\psi_d = 0$), which follows from

$$\begin{aligned}
\underset{\text{(age,area)}}{\text{GM}}[\text{Odds(female,trn1)}] &= \left(\prod_{k=1}^{32}\prod_{g=1}^{10} \frac{\alpha_{sgk}^{(d-1)d}}{1 - \alpha_{sgk}^{(d-1)d}}\right)^{\frac{1}{32\times 10}} \\
&= \left(\prod_{k=1}^{32}\prod_{g=1}^{10} \exp(-\beta_0) \times \exp\left(-\delta_g^{(sd)}\right) \times \exp\left(-\phi_k^{(d)}\right)\right)^{\frac{1}{32\times 10}} \\
&= \left(\exp(-\beta_0)^{32\times 10} \times \exp\left(-\sum_{g=1}^{10}\delta_g^{(sd)}\right)^{32} \times \exp\left(-\sum_{k=1}^{32}\phi_k^{(d)}\right)^{10}\right)^{\frac{1}{32\times 10}} \\
&= \exp(-\beta_0), \qquad \text{(B.2)}
\end{aligned}$$

where "trn" stands for transition. The last equality follows from the sex- and transition-specific sum-to-zero constraints $\sum_{g=1}^{10}\delta_g^{(sd)} = 0$ and $\sum_{k=1}^{32}\phi_k^{(d)} = 0$.

Similarly, the following hold:

$$\underset{\text{(age,area)}}{\text{GM}}[\text{Odds(male,trn1)}] = \exp(-\beta_0) \times \exp(-\gamma_s), \qquad \text{(B.3)}$$

$$\underset{\text{(age,area)}}{\text{GM}}[\text{Odds(female,trn2)}] = \exp(-\beta_0) \times \exp(-\psi_d), \qquad \text{(B.4)}$$

$$\underset{\text{(age,area)}}{\text{GM}}[\text{Odds(male,trn2)}] = \exp(-\beta_0) \times \exp(-\gamma_s) \times \exp(-\psi_d), \qquad \text{(B.5)}$$

where each geometric mean is computed over all age groups and council areas.

The geometric mean of the ratio of two sequences of equal lengths is the same as the ratio of the geometric means of the two sequences. That is, let $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ be two sequences of equal lengths. The geometric mean of the ratio of the two sequences can be computed as follows:

$$\begin{aligned}
\mathrm{GM}\left[\left(\frac{x_1}{y_1}, \frac{x_2}{y_2}, \ldots, \frac{x_n}{y_n}\right)\right] &= \left(\prod_{i=1}^{n} \frac{x_i}{y_i}\right)^{1/n} \\
&= \frac{\left(\prod_{i=1}^{n} x_i\right)^{1/n}}{\left(\prod_{i=1}^{n} y_i\right)^{1/n}} \\
&= \frac{\mathrm{GM}[(x_1, x_2, \ldots, x_n)]}{\mathrm{GM}[(y_1, y_2, \ldots, y_n)]}.
\end{aligned}$$

The same logic applies to odds ratios. That is, the geometric mean of the odds ratios of attrition for males compared to females, over all age groups and council areas, can be computed (using Equations B.2-B.5) as

$$\begin{aligned}
\underset{\text{(age,area)}}{\mathrm{GM}}\left[\frac{\text{Odds(male,trn1)}}{\text{Odds(female,trn1)}}\right] &= \frac{\underset{\text{(age,area)}}{\mathrm{GM}}[\text{Odds(male,trn1)}]}{\underset{\text{(age,area)}}{\mathrm{GM}}[\text{Odds(female,trn1)}]} \\
&= \frac{\exp(-\beta_0) \times \exp(-\gamma_s)}{\exp(-\beta_0)} \\
&= \exp(-\gamma_s)\left(= \underset{\text{(age,area)}}{\mathrm{GM}}\left[\frac{\text{Odds(male,trn2)}}{\text{Odds(female,trn2)}}\right]\right).
\end{aligned}$$

Thus, for both transitions, the odds ratio of attrition if one is male as opposed to female, averaged over all age groups and council areas, reduces to $\exp(-\gamma_s)$.

Similarly, the geometric mean of the odds ratios of attrition for the second transition compared to the first, over all age groups and council areas, can be computed as

$$\underset{\text{(age,area)}}{\mathrm{GM}}\left[\frac{\text{Odds(male,trn2)}}{\text{Odds(male,trn1)}}\right] = \exp(-\psi_d)\left(= \underset{\text{(age,area)}}{\mathrm{GM}}\left[\frac{\text{Odds(female,trn2)}}{\text{Odds(female,trn1)}}\right]\right).$$

Thus, for both sexes, the term $\exp(-\psi_d)$ is the average odds ratio of attrition if one is in the transition from doses 2 to 3 compared to the transition from doses 1

to 2.

An interpretation of the random effects can be found using a similar strategy. For example, for $s =$ female, $d = 2$ (transition from doses 1 to 2), council area k, and age group g_1, the following holds:

$$\begin{aligned}
\frac{\text{Odds(female,trn1,age } g_1, \text{area } k)}{\underset{\text{(age)}}{\text{GM}}[\text{Odds(female,trn1,area } k)]} &= \frac{\exp(-\beta_0) \times \exp\left(-\delta_{g_1}^{(sd)}\right) \times \exp\left(-\phi_k^{(d)}\right)}{\left(\prod_{g=1}^{10} \exp(-\beta_0) \times \exp\left(-\delta_g^{(sd)}\right) \times \exp\left(-\phi_k^{(d)}\right)\right)^{\frac{1}{10}}} \\
&= \frac{\exp(-\beta_0) \times \exp\left(-\delta_{g_1}^{(sd)}\right) \times \exp\left(-\phi_k^{(d)}\right)}{\left(\exp(-\beta_0)^{10} \times \exp\left(-\sum_{g=1}^{10} \delta_g^{(sd)}\right) \times \exp\left(-\phi_k^{(d)}\right)^{10}\right)^{\frac{1}{10}}} \\
&= \frac{\exp(-\beta_0) \times \exp\left(-\delta_{g_1}^{(sd)}\right) \times \exp\left(-\phi_k^{(d)}\right)}{\exp(-\beta_0) \times \exp\left(-\phi_k^{(d)}\right)} \\
&= \exp\left(-\delta_{g_1}^{(sd)}\right),
\end{aligned}$$

where the third equality follows from the sum-to-zero constraint $\sum_{g=1}^{10} \delta_g^{(sd)} = 0$. Generally, for any sex, transition, and council area, it follows that $\exp\left(-\delta_{g_1}^{(sd)}\right)$ is the odds ratio of attrition for age group g_1 compared to the sex- and transition-specific average (geometric mean) odds across all age groups. Similarly, it can be shown that $\exp\left(-\phi_{k_1}^{(d)}\right)$ is the odds ratio of attrition for council area k_1, compared to the geometric mean of the odds over all council areas, where sex, transition, and age group are fixed. Thus, $\exp\left(-\phi_{k_1}^{(sd)}\right)$ is the odds ratio of attrition for council area k_1 compared to the sex- and transition-specific Scottish average.

B.5 Observed vs fitted attrition rates

Figure B.4 shows the fitted attrition rates from Model 1 plotted against the observed attrition rates, where the colour of the point denotes the age group. The dashed line represents the line of equality. Most points fall near the dashed equality line and no unusual patterns are visible, suggesting an appropriate model fit. The deviations of the observed vs fitted attrition rates from the equality line are greatest for the

age group 18 to 29 years, which reflects the overall greater variation in the observed attrition rates in that age group, as seen in Figure 4.1.

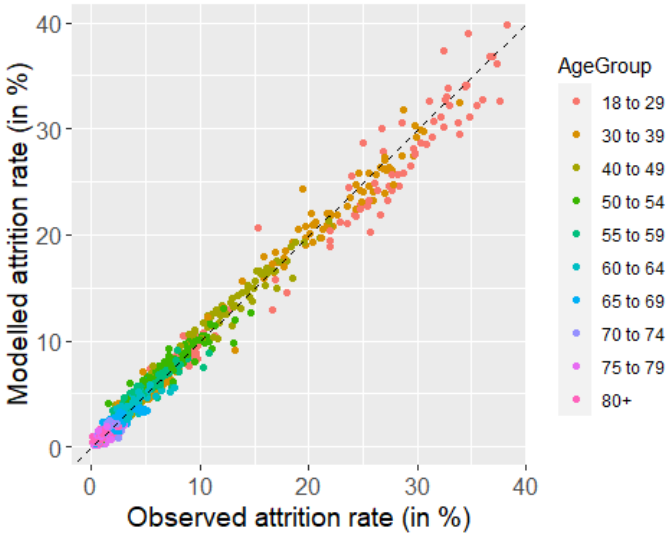


Figure B.4: Observed vs fitted attrition rates.

B.6 Proof of principle simulation

This section provides a proof of principle simulation to check if the parameters from Model 2 can be well estimated, as follows. Equation (4.5) assigns a log-linear model to the odds in favour of continuation of the vaccination program, as

$$\log\left(\frac{1-\alpha_{sgk}^{(d-1)d}}{\alpha_{sgk}^{(d-1)d}}\right) = \beta_0 + \gamma_s + \psi_d + \delta_g^{(sd)} + \phi_k^{(sd)}.$$

Rewriting the above equation gives the modelled attrition rate as

$$\alpha_{sgk}^{(d-1)d} = \frac{\exp\left(-\beta_0 - \gamma_s - \psi_d - \delta_g^{(sd)} - \phi_k^{(sd)}\right)}{1+\exp\left(-\beta_0 - \gamma_s - \psi_d - \delta_g^{(sd)} - \phi_k^{(sd)}\right)}.$$

Thus, assigning values to all β_0, γ_s, ψ_d, $\delta_g^{(sd)}$, and $\phi_k^{(sd)}$ allows the computation of $\alpha_{sgk}^{(d-1)d}$ for all s, g, k, and d. The aim of this proof of principle simulation is to confirm that the proposed method is able to correctly estimate the parameters in the model for data similar to that observed in this study. Hence, here I use the observed data to compute model-free estimates of β_0, γ_s, ψ_d, $\delta_g^{(sd)}$, and $\phi_k^{(sd)}$.

Let Odds(sex s, age g, trn $(d-1)$, area k)$=A_{sgk}^{(d-1)d} \Big/ \left(1-A_{sgk}^{(d-1)d}\right)$ denote the odds in favour of attrition, computed from the observed attrition rate $A_{sgk}^{(d-1)d} = 1-\left(Y_{sgkd} \Big/ Y_{sgk(d-1)}\right)$. From Section B.4, one can obtain the following estimates, and in what follows GM denotes the geometric mean.

- $\hat{\beta}_0 = -\log\left(\underset{\text{(age,area)}}{\text{GM}}[\text{Odds(female,trn1)}]\right)$.
- For γ_s, there are two possible estimates

$$\hat{\gamma}_s^{(1)} = -\log\left(\frac{\underset{\text{(age,area)}}{\text{GM}}[\text{Odds(male,trn1)}]}{\underset{\text{(age,area)}}{\text{GM}}[\text{Odds(female,trn1)}]}\right)$$

  and

$$\hat{\gamma}_s^{(2)} = -\log\left(\frac{\underset{\text{(age,area)}}{\text{GM}}[\text{Odds(male,trn2)}]}{\underset{\text{(age,area)}}{\text{GM}}[\text{Odds(female,trn2)}]}\right),$$

  so I let $\hat{\gamma}_s = (\hat{\gamma}_s^{(1)} + \hat{\gamma}_s^{(2)})/2$.

- Similarly, for ψ_d there are two possible estimates

$$\hat{\psi}_d^{(1)} = -\log\left(\frac{\underset{\text{(age,area)}}{\text{GM}}[\text{Odds(female,trn2)}]}{\underset{\text{(age,area)}}{\text{GM}}[\text{Odds(female,trn1)}]}\right)$$

  and

$$\hat{\psi}_d^{(2)} = -\log\left(\frac{\underset{\text{(age,area)}}{\text{GM}}[\text{Odds(male,trn2)}]}{\underset{\text{(age,area)}}{\text{GM}}[\text{Odds(male,trn1)}]}\right),$$

  so I let $\hat{\psi}_d = (\hat{\psi}_d^{(1)} + \hat{\psi}_d^{(2)})/2$.

- For each age group $g = 1, \ldots, G$, there is one estimate of $\delta_g^{(sd)}$ for each council area. For council areas $k = 1, \ldots, K$, I let

$$\hat{\delta}_g^{(sd)(k)} = -\log\left(\frac{\text{Odds(sex } s, \text{trn } d, \text{age } g, \text{area } k)}{\underset{\text{(age)}}{\text{GM}}[\text{Odds(sex } s, \text{trn } d, \text{area } k)]}\right),$$

  and estimate $\delta_g^{(sd)}$ as the average $\hat{\delta}_g^{(sd)} = \frac{1}{K}\sum_{k=1}^{K} \hat{\delta}_g^{(sd)(k)}$.

- For each council area $k = 1, \ldots, K$, there is one estimate of $\phi_k^{(sd)}$ for each age group. For age groups $g = 1, \ldots, G$, I let

$$\hat{\phi}_k^{(sd)(g)} = -\log\left(\frac{\text{Odds(sex } s, \text{trn } d, \text{age } g, \text{area } k)}{\underset{\text{(area)}}{\text{GM}}[\text{Odds(sex } s, \text{trn } d, \text{age } g)]}\right),$$

and estimate $\phi_k^{(sd)}$ as the average $\hat{\phi}_k = \frac{1}{G}\sum_{g=1}^{G} \hat{\phi}_k^{(sd)(g)}$.

These computations result in the following model-free estimates of the fixed effects:

$$\hat{\beta}_0 = 4.36, \quad \hat{\gamma}_s = -0.09, \quad \hat{\psi}_d = -1.35.$$

The model-free estimates of the random effects can be found in Tables B.2 and B.3.

Table B.1: The true parameter values of the fixed effects with the bias and standard deviation of the estimates, averaged over the 100 simulated datasets.

Effect	Parameter	True parameter value	Bias	Standard deviation
Intercept	β_0	4.36	0.0015	0.0375
Sex	γ_s	-0.09	-0.0015	0.0432
Transition	ψ_d	-1.35	-0.0014	0.0432

Table B.2: True parameter values of the age group random effects with the bias and standard deviation (SD) of the estimates, averaged over the 100 simulated datasets.

Age group	Doses 1 to 2						Doses 2 to 3					
	Female			Male			Female			Male		
	True	Bias	SD	True	Bias	SD	True	Bias	SD	True	Bias	SD
80+	0.71	0.0035	0.03	1.24	0.0027	0.04	1.38	-0.0014	0.02	1.82	0.0027	0.03
75 to 79	1.06	0.0027	0.04	1.25	0.0003	0.05	1.24	0.0006	0.03	1.65	-0.0001	0.03
70 to 74	0.92	0.0054	0.04	1.11	0.0003	0.04	1.17	0.0022	0.02	1.33	-0.0029	0.02
65 to 69	0.77	0.0004	0.03	0.75	-0.0009	0.03	0.73	0.0017	0.02	0.82	0.0003	0.02
60 to 64	0.30	-0.0010	0.03	0.39	-0.0036	0.03	0.37	-0.0004	0.02	0.36	0.0002	0.02
55 to 59	0.09	-0.0038	0.02	0.01	-0.0002	0.02	-0.06	-0.0001	0.02	-0.18	0.0003	0.02
50 to 54	-0.17	-0.0027	0.02	-0.39	0.0012	0.02	-0.34	-0.0000	0.02	-0.59	-0.0007	0.02
40 to 49	-0.69	-0.0016	0.02	-0.95	-0.0005	0.02	-0.95	-0.0014	0.01	-1.21	0.0003	0.02
30 to 39	-1.31	-0.0010	0.02	-1.52	-0.0003	0.02	-1.61	-0.0007	0.01	-1.83	0.0006	0.01
18 to 29	-1.67	-0.0014	0.02	-1.89	0.0003	0.02	-1.91	-0.0009	0.01	-2.18	-0.0000	0.01

Assume the number of people who received a first dose of the vaccine, Y_{sgk1}, are known for each combination of sex, age group, and council area (I use the observed counts from the dataset). For the parameter values above, generate the number of people who received a second dose ($d = 2$) as $Y_{sgk2} \sim \text{Binomial}\left(Y_{sgk1}, 1 - \alpha_{sgk}^{(d-1)d}\right)$. Then, from the generated values $\{Y_{sgk2}\}$, generate the number of people who received

a third dose ($d = 3$) as $Y_{sgk3} \sim \text{Binomial}\left(Y_{sgk2}, 1 - \alpha_{sgk}^{(d-1)d}\right)$. Finally, fit Model 2 to these simulated data. I repeat this procedure 100 times and in each trial I compute the bias of the posterior mean estimate and the posterior standard deviation for each parameter. Table B.1 shows the bias and the standard deviation averages over the 100 trials for the fixed effects. Table B.2 shows these results for the age group random effects and Table B.3 shows those for the council area random effects.

Table B.3: True parameter values of the council area random effects with the bias and standard deviation (SD) of the estimates, averaged over the 100 simulated datasets.

Council area	Doses 1 to 2						Doses 2 to 3					
	Female			Male			Female			Male		
	True	Bias	SD	True	Bias	SD	True	Bias	SD	True	Bias	SD
West Lothian	0.02	0.00	0.04	-0.05	-0.00	0.04	-0.24	0.00	0.04	-0.13	0.00	0.04
West Dunbartonshire	-0.19	0.00	0.05	-0.19	0.01	0.05	-0.45	0.00	0.04	-0.34	0.00	0.04
Stirling	0.06	0.00	0.05	0.25	-0.01	0.05	0.24	0.00	0.04	0.17	0.00	0.04
South Lanarkshire	-0.07	-0.00	0.04	-0.06	0.00	0.04	-0.03	-0.00	0.04	-0.02	0.00	0.04
South Ayrshire	-0.05	0.01	0.05	-0.09	-0.00	0.04	0.26	-0.00	0.04	0.44	-0.00	0.04
Shetland Islands	0.35	-0.01	0.09	0.58	-0.04	0.08	0.82	-0.01	0.06	0.32	0.00	0.05
Scottish Borders	-0.03	0.00	0.05	0.11	-0.00	0.05	0.08	-0.00	0.04	0.01	0.00	0.04
Renfrewshire	-0.05	0.00	0.04	-0.04	-0.00	0.04	-0.06	0.00	0.04	-0.10	0.00	0.04
Perth and Kinross	0.23	-0.01	0.05	0.18	0.00	0.04	0.17	0.00	0.04	0.12	0.00	0.04
Orkney Islands	0.04	0.01	0.08	0.06	0.01	0.07	0.73	-0.02	0.06	0.53	-0.01	0.05
North Lanarkshire	-0.16	-0.00	0.04	-0.21	0.00	0.04	-0.31	-0.00	0.04	-0.25	0.00	0.04
North Ayrshire	-0.21	0.00	0.04	-0.06	0.00	0.04	-0.16	0.00	0.04	-0.26	0.00	0.04
Na h-Eileanan Siar	0.07	0.00	0.08	-0.25	0.01	0.06	-0.02	0.01	0.05	0.06	0.00	0.05
Moray	0.06	-0.00	0.05	-0.10	0.00	0.05	-0.04	0.00	0.04	0.08	-0.00	0.04
Midlothian	0.03	-0.00	0.05	0.12	0.00	0.05	0.05	0.00	0.04	0.02	0.00	0.04
Inverclyde	-0.02	-0.00	0.05	-0.05	0.01	0.05	0.04	0.00	0.04	0.03	0.00	0.04
Highland	0.04	0.00	0.04	-0.07	0.00	0.04	-0.05	-0.00	0.04	-0.07	0.00	0.04
Glasgow City	-0.54	0.00	0.04	-0.56	-0.00	0.04	-0.65	0.00	0.04	-0.57	0.00	0.04
Fife	-0.18	0.00	0.04	-0.14	0.00	0.04	-0.10	-0.00	0.04	-0.05	0.00	0.04
Falkirk	0.16	-0.00	0.04	0.22	-0.00	0.04	-0.20	0.00	0.04	-0.23	0.00	0.04
East Renfrewshire	0.27	-0.00	0.05	0.27	-0.00	0.05	0.24	-0.00	0.04	0.48	-0.00	0.04
East Lothian	0.19	-0.00	0.05	0.12	-0.00	0.05	0.20	-0.00	0.04	0.22	-0.00	0.04
East Dunbartonshire	0.43	-0.01	0.05	0.45	-0.01	0.05	0.34	0.00	0.04	0.50	-0.00	0.04
East Ayrshire	-0.02	0.00	0.05	0.07	-0.00	0.04	-0.09	-0.00	0.04	-0.13	0.00	0.04
Dundee City	-0.31	0.00	0.04	-0.31	0.00	0.04	-0.52	0.00	0.04	-0.48	0.00	0.04
Dumfries and Galloway	-0.20	0.00	0.04	-0.25	0.00	0.04	0.23	-0.00	0.04	0.19	0.00	0.04
Clackmannanshire	-0.00	0.00	0.06	0.08	-0.01	0.05	-0.15	0.00	0.04	-0.19	0.00	0.04
City of Edinburgh	-0.20	-0.00	0.04	-0.17	0.00	0.04	-0.07	0.00	0.04	-0.14	0.00	0.04
Argyll and Bute	-0.16	0.01	0.05	-0.27	0.00	0.04	-0.12	-0.00	0.04	-0.08	0.00	0.04
Angus	0.25	-0.00	0.05	0.22	-0.00	0.05	0.00	-0.00	0.04	-0.03	-0.00	0.04
Aberdeenshire	0.23	-0.00	0.04	0.15	-0.00	0.04	0.07	-0.00	0.04	0.12	0.00	0.04
Aberdeen City	-0.03	0.00	0.04	-0.02	0.00	0.04	-0.23	0.00	0.04	-0.22	0.00	0.04

The tables show that over the 100 simulated datasets, the bias tends to be very close to zero for all parameter estimates, which suggests that the model is able to

produce good estimates of the true parameter values. The posterior standard deviations appear to be relatively small compared to the true values, suggesting good estimation performance.

B.7 Sensitivity analysis for hyperprior specifications

The hierarchical model requires specification of parameter values for the prior and hyperprior distributions, and the choice of hyperparameters for the hyperprior distributions is not straightforward. The results for the best fitting model were obtained using `R-INLA`'s default hyperprior specifications for the penalised complexity (PC) prior distributions of the precision parameters τ_δ and τ_ϕ, as well as the mixing parameters ρ_δ and ρ_ϕ. This section provides a sensitivity analysis of the choice of parameter values for the corresponding hyperprior distributions to see if they substantially impact the posterior distributions of these parameters and, therefore, the estimated effects overall.

The precision parameters τ_δ and τ_ϕ are assigned type-2 Gumbel prior distributions of the form $\pi(\tau) = \frac{\theta}{2}\tau^{-3/2}\exp(-\theta\tau^{-1/2})$, for $\tau \in \{\tau_\delta, \tau_\phi\}$, which were proposed by Simpson et al. (2017). Choosing the value of the parameter θ can be aided by considering the probability statement $P(1/\sqrt{\tau} > U) = \alpha$ (Riebler et al., 2016), and I use `R-INLA`'s default values $U = 1$ and $\alpha = 0.01$. The mixing parameters ρ_δ and ρ_ϕ are assigned another PC prior distribution, where the base model is obtained for $\rho = 0$. The Kullback-Leibler divergence (KLD, Kullback and Leibler (1951)) for between the base model and corresponding flexible BYM2 model is used to compute the distance scale $d(\rho)$ (recall Section 2.3.1.1), which is assigned an exponential prior distribution with parameter λ. In contrast to the precision parameter τ, the PC prior for ρ is not available in closed form but can be computed in `R-INLA` for the application-specific structure matrix (see Riebler et al., 2016 for more details). The probability statement $P(\rho < U) = \alpha$ can be used to determine a reasonable value of λ, and I use `R-INLA`'s default values $U = 0.5$ and $\alpha = 0.5$.

Table B.4 contains the means and 95% credible intervals for the posterior distri-

Table B.4: Mean estimates and 95% credible intervals (in parantheses) for the posterior distributions of the hyperpriors used in Model 1, for different hyperparameter specifications.

Prior name	Hyperparameter settings	τ_δ	ρ_δ	τ_ϕ	ρ_ϕ
Default	$U_\tau = 1$, $\alpha_\tau = 0.01$	3.907	1.000	19.873	0.414
	$U_\rho = 0.5$, $\alpha_\rho = 0.5$	(2.513,5.230)	(0.998,1.000)	(13.783,27.210)	(0.232,0.700)
Prior 1	$U_\tau = 1$, $\alpha_\tau = 0.01$	4.322	1.000	21.384	0.293
	$U_\rho = 0.5$, $\alpha_\rho = 2/3$	(2.660,6.505)	(0.997,1.000)	(15.423,29.540)	(0.092,0.549)
Prior 2	$U_\tau = 1$, $\alpha_\tau = 0.05$	4.225	0.991	20.490	0.324
	$U_\rho = 0.5$, $\alpha_\rho = 0.5$	(2.571,6.401)	(0.917,1.000)	(13.832,28.836)	(0.102,0.627)
Prior 3	$U_\tau = 1$, $\alpha_\tau = 0.05$	4.239	0.998	20.523	0.320
	$U_\rho = 0.5$, $\alpha_\rho = 2/3$	(2.586,6.429)	(0.993,1.000)	(13.834,28.978)	(0.070,0.691)

butions of τ and ρ for the age group effects δ and council area effects ϕ for different prior specifications of Model 1. The first row in the table corresponds to the model for which the resulting estimates are presented in Section 4.4. The other prior specifications are considered for sensitivity analysis.

The main takeaway from the table is that changing the prior specifications does not substantially impact the posterior distributions, as the mean estimates and 95% credible intervals of each hyperparameter are fairly similar for the different prior specifications. The precision of the variance parameter τ_δ of the age group effects with prior specification $P(1/\sqrt{\tau} > 1) = 0.01$ (which was used in this study) tends to be slightly lower than that of the other prior specifications. The minor differences in posterior estimates seem negligible. These differences are even more subtle for the variance parameter τ_ϕ. For the mixing parameter ρ_δ, the posterior mean estimates and 95% credible intervals are all very close to 1, suggesting that the structured component almost entirely explains the variation in the age group effects, and this result is not sensitive to the hyperparameter specifications. The means of the posterior distributions for the mixing parameter ρ_ϕ are all moderate, falling between 0.29 and 0.42, and the 95% credible intervals are fairly wide, with upper bounds as high as 0.70.

Lastly, Figures B.5 shows the posterior mean estimates of the sex- and transition-specific odds ratios of attrition for each age group, compared to the average over all age groups, i.e. $\exp(-\delta_g^{(sd)})$, for the hyperprior specifications presented in Table B.4; similarly, Figure B.6 shows the posterior mean estimates of the transition-specific odds ratios of attrition for each council area, compared to the Scottish average, i.e.

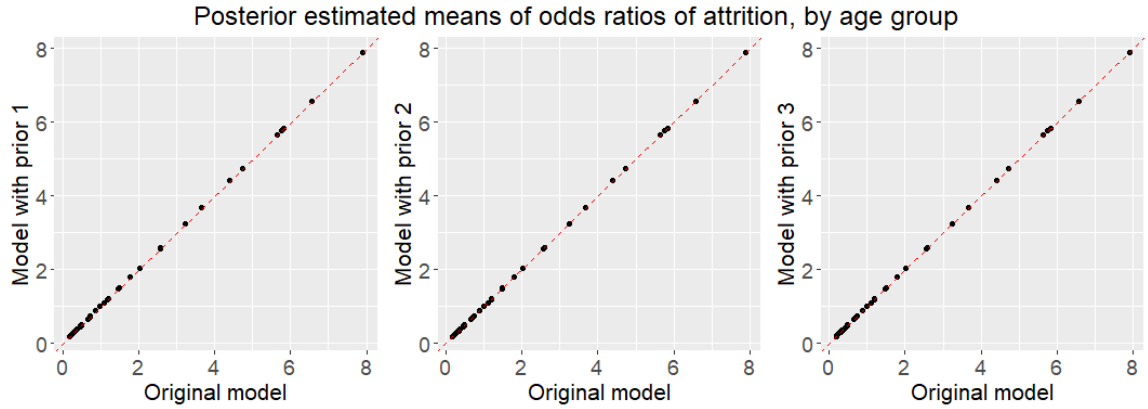


Figure B.5: Posterior mean estimates of the sex- and transition-specific odds ratios of attrition for each age group, compared to the average over all age groups for different hyperprior specifications.

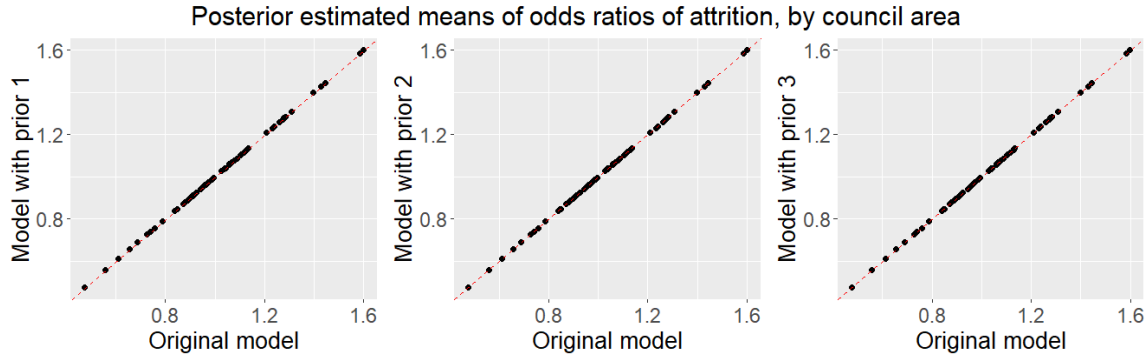


Figure B.6: Posterior mean estimates of the transition-specific odds ratios of attrition for each council area, compared to the Scottish average for different hyperprior specifications.

$\exp(-\phi_k^{(d)})$. The figures confirm that the choice of the prior specification has almost no impact on the resulting modelled attrition rates, as the modelled odds ratios of attrition obtained from the models with alternative hyperprior specifications are almost identical to those presented in Section 4.4.

B.8 Sensitivity analysis for the random effects specifications

The BYM2 models assigned to the age group and council area random effects require a structure for the correlated random effects. The proposed model uses a binary adjacency neighbourhood matrix $\boldsymbol{M}$ for the correlated council area random effects $\boldsymbol{u}_\phi^{*(sd)}$. Here, $m_{ij} = 1$ if areas i and j share a border and $m_{ij} = 0$, otherwise. Additionally, I created artificial links (setting $m_{ij} = 1$) between the islands and their nearest land areas. However, these artificially linked council areas might be less

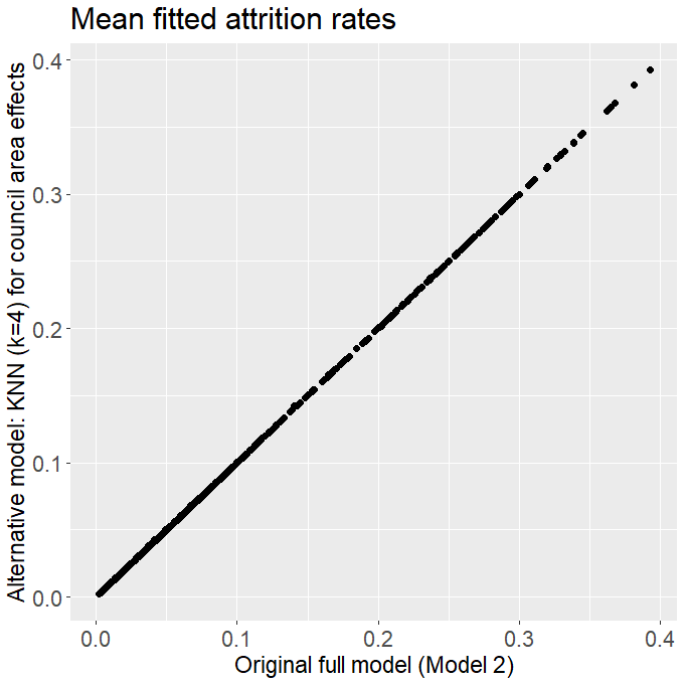


Figure B.7: Mean fitted attrition rates of the original Model 2 and the alternative that uses a k nearest neighbour matrix with $k = 4$ for the scaled correlated council area random effects $\boldsymbol{u}_{\phi}^{*(sd)}$.

similar to each other than the council areas that share a border. To check if the results are sensitive to the choice of structure matrix or the artificially created links, I fit an alternative version of Model 2 using a k-nearest neighbour (KNN) matrix $\boldsymbol{M}^*$. Since the border-sharing rule results in an average of 3.94 neighbours per area, I let $k = 4$ and treat two areas as neighbours if one of them is amongst the four nearest neighbours (as measured by the Euclidean distance between the polygon's centroids) of the other. I symmetrise this neighbourhood structure so that if area i is amongst the four nearest neighbours of area j, then area j is also labelled as a neighbour of area i even if it is not amongst its four nearest neighbours. The resulting mean fitted attrition rates are very similar to those of Model 2, as shown in Figure B.7. The mean absolute difference between the mean fitted attrition rates from the two models is approximately 0.000039, so the results do not appear to be sensitive to the choice of neighbourhood matrix used in the CAR model. On the same note, the results do not seem to be strongly affected by the artificially linked islands.

In the BYM2 prior of the age group effects, the neighbourhood structure assigned to the scaled correlated age group random effects $\boldsymbol{u}_{\delta}^{*(sd)}$ implies that only neighbouring age groups are partially correlated. I check the sensitivity of this assumption by fitting an alternative model that assigns an autoregressive (AR) prior of order 2 to the age

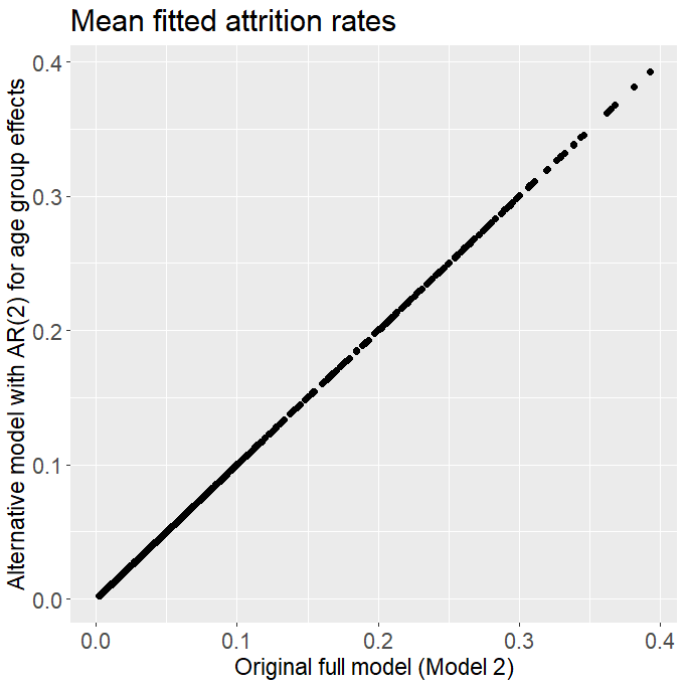


Figure B.8: Mean fitted attrition rates of the original Model 2 and the alternative that assigns an AR(2) prior to the age group random effects $\delta_g^{(sd)}$.

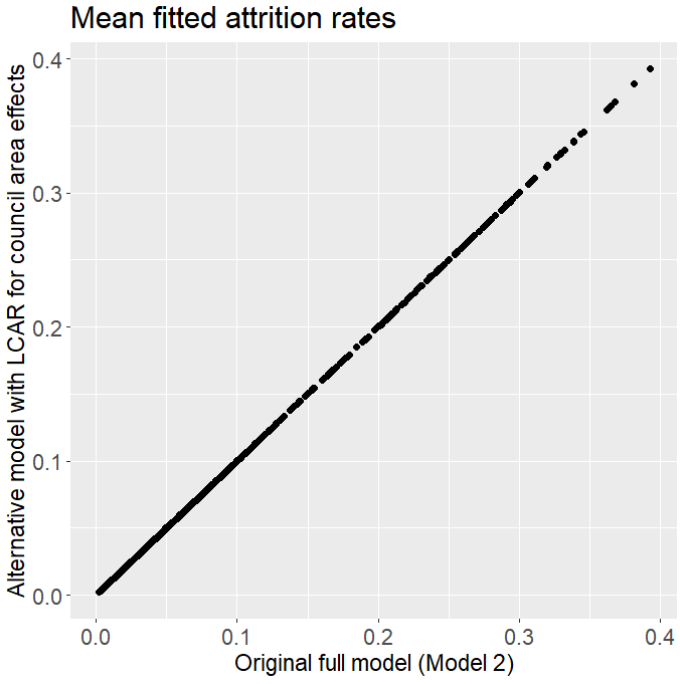


Figure B.9: Mean fitted attrition rates of the original Model 2 and the alternative that assigns a Leroux CAR prior to the council area random effects $\phi_k^{(sd)}$.

group effects $\delta_g^{(sd)}$, so that the effects are partially correlated with neighbours within two age groups distance. The resulting mean fitted attrition rates are very similar to those of Model 2, as shown in Figure B.8, with a mean absolute difference of approximately 0.000054. I could also consider an alternative prior for the council area random effects, moving away from the BYM2 prior I proposed in Chapter 4.

Since the model should flexibly estimate the strength of the spatial autocorrelation in the data, a good alternative is the Leroux CAR (LCAR) prior (Leroux et al., 2000, recall Section 2.6.2). Replacing the BYM2 model, I assign to the council area random effects $\phi_k^{(sd)}$ an LCAR prior that uses the same neighbourhood matrix $\boldsymbol{M}$. This results in the mean fitted attrition rates presented in Figure B.9, which are plotted against the corresponding values from the original Model 2. The mean absolute difference between the fitted attrition rates from the alternative model and the original Model 2 is approximately 0.000030, so the results do not appear to be sensitive to the choice of prior assigned to the council area random effects.

B.9 Posterior means and credible intervals of the odds ratios, by council area

Table B.5 contains posterior mean estimates and 95% credible intervals of the odds ratio of attrition for each council area compared to the transition-specific Scottish averages, corresponding to Figure 4.4 in Chapter 4. The table confirms that over both transitions, Glasgow City, West Dunbartonshire, and Dundee City are the council areas with the highest odds of attrition in Scotland. The council areas with the lowest odds are the Shetland Islands, Orkney Islands, East Dunbartonshire, and East Renfrewshire. Further, the table shows that the 95% credible intervals are wholly above one (indicated by the entries in **bold**) in both transitions for Dundee City, East Ayrshire, Glasgow City, North Ayrshire, North Lanarkshire, Renfrewshire, and West Dunbartonshire. Therefore, in these areas, in both transitions the odds in favour of attrition are significantly larger than the transition-specific Scottish national averages. The 95% credible intervals are wholly below one (indicated by the entries in *italic*) in both transitions for East Dunbartonshire, East Lothian, East Renfrewshire, Western Isles Na h-Eileanan Siar, Orkney Islands, Scottish Borders, Shetland Islands, and Stirling. Therefore, in these areas, in both transitions the odds in favour of attrition are significantly smaller than the transition-specific Scottish national averages.

Table B.6 contains posterior mean estimates and 95% credible intervals of the odds ratio of attrition for each council area compared to the sex- and transition-specific

Table B.5: Posterior mean estimates and 95% credible intervals of the odds ratios of attrition for each council area, compared to the transition-specific average odds across all council areas.

Council area	Doses 1 to 2	Doses 2 to 3
West Lothian	1.03 (0.96,1.10)	**1.27 (1.19,1.36)**
West Dunbartonshire	**1.28 (1.19,1.39)**	**1.44 (1.35,1.55)**
Stirling	*0.91 (0.84,0.98)*	*0.84 (0.78,0.90)*
South Lanarkshire	**1.09 (1.01,1.16)**	1.04 (0.98,1.11)
South Ayrshire	0.99 (0.92,1.07)	0.94 (0.88,1.01)
Shetland Islands	*0.48 (0.41,0.55)*	*0.61 (0.56,0.67)*
Scottish Borders	*0.91 (0.84,0.99)*	*0.91 (0.85,0.97)*
Renfrewshire	**1.10 (1.03,1.18)**	**1.13 (1.06,1.21)**
Perth and Kinross	0.93 (0.86,1.00)	0.99 (0.92,1.06)
Orkney Islands	*0.69 (0.61,0.78)*	*0.56 (0.51,0.61)*
North Lanarkshire	**1.23 (1.15,1.32)**	**1.28 (1.20,1.36)**
North Ayrshire	**1.21 (1.12,1.30)**	**1.24 (1.16,1.33)**
Na h-Eileanan Siar	*0.85 (0.76,0.95)*	*0.88 (0.81,0.95)*
Moray	1.04 (0.96,1.12)	0.99 (0.92,1.06)
Midlothian	0.95 (0.88,1.03)	**1.07 (1.00,1.14)**
Inverclyde	1.06 (0.98,1.15)	1.06 (0.99,1.14)
Highland	1.04 (0.97,1.12)	1.04 (0.97,1.11)
Glasgow City	**1.59 (1.48,1.69)**	**1.40 (1.31,1.49)**
Fife	**1.26 (1.18,1.35)**	1.06 (0.99,1.13)
Falkirk	*0.89 (0.83,0.96)*	**1.08 (1.01,1.15)**
East Renfrewshire	*0.74 (0.68,0.80)*	*0.79 (0.73,0.84)*
East Lothian	*0.90 (0.83,0.97)*	*0.87 (0.81,0.93)*
East Dunbartonshire	*0.66 (0.61,0.71)*	*0.76 (0.71,0.81)*
East Ayrshire	**1.11 (1.03,1.19)**	**1.13 (1.06,1.21)**
Dundee City	**1.43 (1.33,1.53)**	**1.31 (1.22,1.40)**
Dumfries and Galloway	1.07 (0.99,1.15)	*0.91 (0.85,0.98)*
Clackmannanshire	1.03 (0.95,1.13)	**1.13 (1.06,1.22)**
City of Edinburgh	**1.13 (1.06,1.21)**	*0.85 (0.79,0.90)*
Argyll and Bute	**1.60 (1.48,1.72)**	0.96 (0.90,1.03)
Angus	0.97 (0.90,1.05)	1.06 (0.99,1.14)
Aberdeenshire	*0.73 (0.68,0.78)*	0.96 (0.90,1.03)
Aberdeen City	1.06 (0.99,1.14)	**1.12 (1.04,1.19)**

Scottish averages, corresponding to Figure 4.5 in Chapter 4. The table confirms that over both sexes and transitions, Glasgow City, West Dunbartonshire, and Dundee City are the council areas with the highest odds of attrition in Scotland. The council areas with the lowest odds are the Shetland Islands, Orkney Islands, East Dunbartonshire, and East Renfrewshire. Further, the table shows that the 95% credible intervals are wholly above one for each sex and transition for Dundee City, East Ayrshire, Glasgow City, North Ayrshire, North Lanarkshire, and West Dunbartonshire. Therefore, in these areas, for each sex and transition the odds in favour of attrition are significantly larger than the sex- and transition-specific Scottish national aver-

ages. The 95% credible intervals are wholly below one for each sex and transition for East Dunbartonshire, East Lothian, East Renfrewshire, Orkney Islands, and Shetland Islands. Therefore, in these areas, for each sex and transition the odds in favour of attrition are significantly smaller than the sex- and transition-specific Scottish national averages.

Table B.6: Posterior mean estimates and 95% credible intervals of the odds ratios of attrition for each council area, compared to the sex- and transition-specific average odds across all council areas.

Council area	Doses 1 to 2		Doses 2 to 3	
	Female	Male	Female	Male
West Lothian	1.02 (0.95,1.10)	1.03 (0.96,1.11)	**1.30 (1.22,1.38)**	**1.25 (1.17,1.32)**
West Dunbartonshire	**1.24 (1.14,1.35)**	**1.32 (1.22,1.42)**	**1.49 (1.39,1.59)**	**1.40 (1.31,1.50)**
Stirling	0.93 (0.86,1.02)	*0.90 (0.82,0.97)*	*0.83 (0.78,0.89)*	*0.84 (0.79,0.90)*
South Lanarkshire	**1.11 (1.04,1.18)**	**1.07 (1.01,1.14)**	**1.07 (1.01,1.14)**	1.02 (0.96,1.08)
South Ayrshire	1.01 (0.93,1.10)	0.98 (0.91,1.06)	0.92 (0.86,0.98)	0.97 (0.90,1.03)
Shetland Islands	*0.53 (0.44,0.63)*	*0.46 (0.38,0.54)*	*0.54 (0.48,0.60)*	*0.67 (0.61,0.74)*
Scottish Borders	0.94 (0.86,1.02)	*0.90 (0.83,0.97)*	*0.89 (0.83,0.95)*	*0.92 (0.86,0.98)*
Renfrewshire	1.07 (0.99,1.15)	**1.14 (1.06,1.22)**	**1.15 (1.08,1.22)**	**1.11 (1.04,1.18)**
Perth and Kinross	*0.88 (0.81,0.95)*	0.97 (0.90,1.04)	0.95 (0.89,1.01)	1.02 (0.96,1.09)
Orkney Islands	*0.72 (0.61,0.84)*	*0.67 (0.58,0.78)*	*0.53 (0.48,0.59)*	*0.59 (0.53,0.65)*
North Lanarkshire	**1.25 (1.17,1.33)**	**1.22 (1.15,1.30)**	**1.33 (1.25,1.41)**	**1.23 (1.16,1.31)**
North Ayrshire	**1.30 (1.21,1.40)**	**1.13 (1.05,1.22)**	**1.24 (1.17,1.33)**	**1.24 (1.16,1.32)**
Na h-Eileanan Siar	0.88 (0.76,1.01)	*0.83 (0.73,0.95)*	*0.87 (0.79,0.95)*	*0.89 (0.82,0.97)*
Moray	0.94 (0.86,1.02)	**1.11 (1.03,1.21)**	1.06 (0.99,1.13)	*0.93 (0.87,0.99)*
Midlothian	0.98 (0.90,1.07)	0.93 (0.86,1.01)	**1.07 (1.00,1.14)**	1.07 (1.00,1.14)
Inverclyde	1.04 (0.95,1.14)	1.08 (0.99,1.17)	**1.08 (1.01,1.16)**	1.05 (0.98,1.13)
Highland	1.00 (0.93,1.08)	1.07 (1.00,1.15)	1.02 (0.96,1.09)	1.05 (0.99,1.12)
Glasgow City	**1.66 (1.57,1.77)**	**1.52 (1.43,1.62)**	**1.46 (1.38,1.55)**	**1.34 (1.26,1.42)**
Fife	**1.29 (1.21,1.38)**	**1.23 (1.16,1.32)**	1.07 (1.00,1.13)	1.05 (0.99,1.12)
Falkirk	*0.87 (0.81,0.94)*	*0.91 (0.84,0.98)*	**1.10 (1.03,1.17)**	1.06 (1.00,1.13)
East Renfrewshire	*0.76 (0.69,0.83)*	*0.73 (0.67,0.80)*	*0.82 (0.76,0.87)*	*0.77 (0.72,0.82)*
East Lothian	*0.90 (0.83,0.98)*	*0.90 (0.83,0.97)*	*0.86 (0.80,0.92)*	*0.88 (0.83,0.94)*
East Dunbartonshire	*0.66 (0.60,0.72)*	*0.67 (0.61,0.73)*	*0.77 (0.72,0.83)*	*0.75 (0.70,0.80)*
East Ayrshire	**1.12 (1.04,1.21)**	**1.09 (1.01,1.18)**	**1.12 (1.05,1.20)**	**1.14 (1.07,1.22)**
Dundee City	**1.48 (1.38,1.59)**	**1.38 (1.29,1.48)**	**1.35 (1.27,1.44)**	**1.26 (1.19,1.34)**
Dumfries and Galloway	1.06 (0.98,1.15)	1.08 (1.00,1.16)	*0.86 (0.81,0.92)*	0.96 (0.90,1.02)
Clackmannanshire	1.09 (0.98,1.20)	0.99 (0.90,1.09)	**1.13 (1.05,1.22)**	**1.14 (1.06,1.22)**
City of Edinburgh	**1.21 (1.14,1.29)**	1.07 (1.00,1.14)	*0.81 (0.77,0.86)*	*0.88 (0.83,0.93)*
Argyll and Bute	**1.15 (1.05,1.25)**	**1.93 (1.80,2.08)**	1.05 (0.98,1.13)	*0.89 (0.83,0.96)*
Angus	0.92 (0.84,1.00)	1.02 (0.95,1.11)	1.05 (0.98,1.12)	**1.08 (1.01,1.15)**
Aberdeenshire	*0.73 (0.68,0.79)*	*0.73 (0.68,0.78)*	0.96 (0.90,1.02)	0.96 (0.91,1.02)
Aberdeen City	1.06 (0.99,1.13)	1.06 (0.99,1.14)	**1.12 (1.06,1.20)**	**1.11 (1.04,1.18)**

B.10 The relationship between attrition rates and other variables

In Chapter 4, I analyse how the trends in attrition rates for the second and third doses of the COVID-19 vaccine differ by age, sex, and council area in Scotland. I chose not to include other covariates in the model to analyse their relationships with attrition rates, as these covariate effects could be impacted by ecological bias (Wakefield and Salway, 2001) due to the data being aggregated to a relatively high council area level. Further, there are only 32 council areas, so there is a risk that any regression relationship will be badly estimated. However, for completeness and with reference to the caveats above, here I briefly examine covariate effects. The covariates I include are a measure of social inequality and population density. The Scottish Government (2020c) provides indicator data used to compute the Scottish Index of Multiple Deprivation (SIMD) at a Data Zone level. One of the main indicators for the SIMD is income deprivation, so I include the income deprivation rate as a covariate in the model. Since the study is at the council area level, I take population-weighted averages of the Data Zone level data. I also include the log-transformed population density (the number of people per square kilometre) as an additional covariate in the model. I apply a log transformation to the population densities because there are a few areas with very high population densities.

I refit the model with the additional covariates, and Figure B.10 shows the mean fitted attrition rates of the new model compared to the mean fitted attrition rates of the original Model 2. The resulting mean fitted attrition rates are almost identical, with a mean absolute difference of approximately 0.0000842, which shows that including the additional covariates has little impact on the results. The fitted parameter values suggest that, as the income deprivation rate increases by one standard deviation, the sex- and transition-specific average odds ratio in favour of attrition over all age groups and council areas increases by a factor of 8, with a posterior mean of 8.18 and a 95% credible interval of (5.74,11.58). Therefore, a higher income deprivation rate appears to be associated with a higher odds ratio in favour of attrition, mean-

ing that on average, those with greater income deprivation are more likely to quit the vaccination programme. The fitted parameter values suggest that increasing the log-population density by one standard deviation has very little impact on the sex- and transition-specific average odds ratio over all age groups and council areas, with a posterior mean of 1.00 and a 95% credible interval of (0.9998,1.0001). However, as previously described with only 32 council areas, it is not clear how well the model can estimate the relationship between the attrition rates and the income deprivation rate and population density covariates.

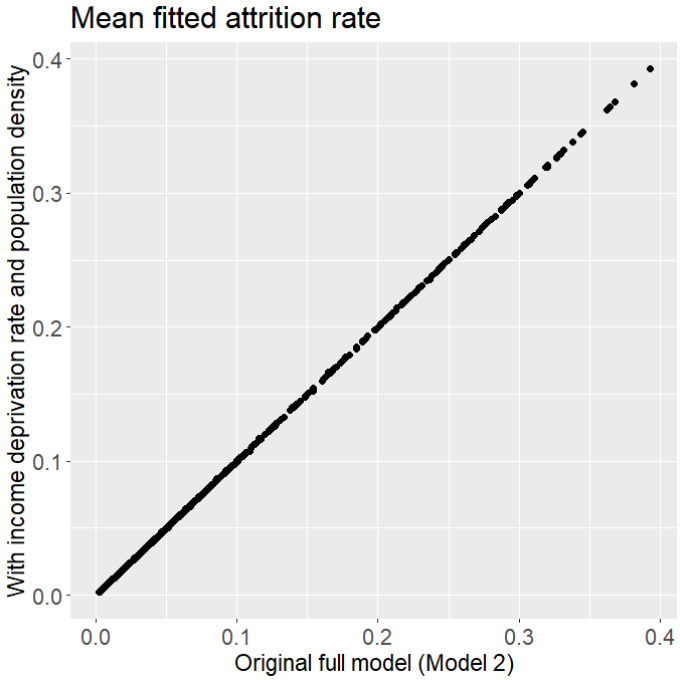


Figure B.10: Mean fitted attrition rates of the original Model 2 and the alternative model with the income deprivation rate and log transformed population density as additional covariates.

References

Adam, N. R., Janeja, V. P., and Atluri, V. (2004). Neighborhood based detection of anomalies in high dimensional spatio-temporal sensor datasets. *Proceedings of the 2004 ACM symposium on Applied computing*, pages 576–583. 137

Aggarwal, C. C. (2017). *Outlier Analysis*. Springer International Publishing. 145

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723. 27

Al-Zoughool, M., Oraby, T., Vainio, H., Gasana, J., Longenecker, J., Ali, W. A., AlSeaidan, M., Elsaadany, S., and Tyshenko, M. G. (2022). Using a stochastic continuous-time Markov chain model to examine alternative timing and duration of the COVID-19 lockdown in Kuwait: What can be done now? *Archives of Public Health*, 80:22. 92

Alderson, M. (1983). *An Introduction to Epidemiology*. Macmillan Education UK. 2

Anderson, E. (1936). The species problem in iris. *Annals of the Missouri Botanical Garden*, 23:457. 51

Andrews, N., Stowe, J., Kirsebom, F., Toffa, S., Sachdeva, R., Gower, C., Ramsay, M., and Bernal, J. L. (2022a). Effectiveness of COVID-19 booster vaccines against COVID-19-related symptoms, hospitalization and death in England. *Nature Medicine*, 28:831–837. 98

Andrews, N., Tessier, E., Stowe, J., Gower, C., Kirsebom, F., Simmons, R., Gallagher, E., Thelwall, S., Groves, N., Dabrera, G., Myers, R., Campbell, C. N., Amirthalingam, G., Edmunds, M., Zambon, M., Brown, K., Hopkins, S., Chand,

M., Ladhani, S. N., Ramsay, M., and Bernal, J. L. (2022b). Duration of protection against mild and severe disease by COVID-19 vaccines. *New England Journal of Medicine*, 386:340–350. 98

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27:93–115. 31, 137, 157

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC. 14

Bashir, M. F., MA, B., and Shahzad, L. (2020). A brief review of socio-economic and environmental impact of COVID-19. *Air Quality, Atmosphere & Health*, 13:1403–1409. 57

Batty, M., Murcio, R., Iacopini, I., Vanhoof, M., and Milton, R. (2021). London in lockdown: Mobility in the pandemic city. In: COVID-19 pandemic, geospatial information, and community resilience. *CRC Press*, pages 229–244. 82

Bayerlein, M. (2024). Regional health care in the EU. *SWP Research Paper*. 192

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418. 12

Bernal, J. L., Andrews, N., Gower, C., Gallagher, E., Simmons, R., Thelwall, S., Stowe, J., Tessier, E., Groves, N., Dabrera, G., Myers, R., Campbell, C. N., Amirthalingam, G., Edmunds, M., Zambon, M., Brown, K. E., Hopkins, S., Chand, M., and Ramsay, M. (2021). Effectiveness of COVID-19 vaccines against the B.1.617.2 (Delta) variant. *New England Journal of Medicine*, 385:585–594. 97

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., and Songini, M. (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*, 14:2433–2443. 38

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–20. 3, 34

Birant, D. and Kut, A. (2006). Spatio-temporal outlier detection in large databases. *Journal of Computing and Information Technology*, 14:291. 137

Biswas, M., Rahaman, S., Biswas, T. K., Haque, Z., and Ibrahim, B. (2021). Association of sex, age, and comorbidities with mortality in COVID-19 patients: A systematic review and meta-analysis. *Intervirology*, 64:36–47. 66

Boehmke, B. and Greenwell, B. (2019). *Hands-On Machine Learning with R*. Chapman and Hall/CRC. 155

Campello, R. J. G. B., Kröger, P., Sander, J., and Zimek, A. (2020). Density-based clustering. *WIREs Data Mining and Knowledge Discovery*, 10. 136

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury advanced series in statistics and decision sciences. Thomson Learning. 16

CDC (2014). Vaccine testing and the approval process. https://www.cdc.gov/vaccines/basics/test-approve.html. Accessed: 2022-05-27. 97

Chandola, V., Banerjee, A., and Kumar, V. (2009). Outlier detection : A survey. *ACM Computing Surveys*, 41. 135

Chatfield, C. (2003). *The analysis of time series: An introduction, sixth edition*. Chapman & Hall/CRC Texts in Statistical Science. Routledge, 2000 Corporate Blvd N.W. Boca Raton, FL 33431, USA. ISBN 978-1-5848-8317-3. 31

Chawla, S. and Sun, P. (2006). SLOM: A new measure for local spatial outliers. *Knowledge and Information Systems*, 9:412–429. 137

Chen, D., Lu, C.-T., Kou, Y., and Chen, F. (2008). On detecting spatial outliers. *GeoInformatica*, 12:455–475. 137

Cheng, T. and Li, Z. (2006). A multiscale approach for spatio-temporal outlier detection. *Transactions in GIS*, 10:253–263. 137

Coccia, M. (2021). The relation between length of lockdown, numbers of infected people and deaths of COVID-19, and economic growth of countries: Lessons learned

to cope with future pandemics similar to COVID-19 and to constrain the deterioration of economic system. *Science of The Total Environment*, 775:145801. 59, 61

Conyon, M. J., He, L., and Thomsen, S. (2020). Lockdowns and COVID-19 deaths in Scandinavia. *SSRN Electronic Journal*. 59

Corpas-Burgos, F. and Martinez-Beneito, M. A. (2020). On the use of adaptive spatial weight matrices from disease mapping multivariate analyses. *Stochastic Environmental Research and Risk Assessment*, 34:531–544. 134

Costanzo, M., Giglio, M. A. R. D., and Roviello, G. N. (2022). Anti-coronavirus vaccines: Past investigations on SARS-CoV-1 and MERS-CoV, the approved vaccines from BioNTech/Pfizer, Moderna, Oxford/AstraZeneca and others under development against SARSCoV- 2 infection. *Current Medicinal Chemistry*, 29:4–18. 97

Cox, L. A. (2021). Toward practical causal epidemiology. *Global Epidemiology*, 3:100065. 191

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley. 23, 29, 31

Davies, N. G., Barnard, R. C., Jarvis, C. I., Russell, T. W., Semple, M. G., Jit, M., and Edmunds, W. J. (2021). Association of tiered restrictions and a second lockdown with COVID-19 deaths and hospital admissions in England: A modelling study. *The Lancet Infectious Diseases*, 21:482–492. 61, 78, 188

Davies, N. G., Kucharski, A. J., Eggo, R. M., Gimma, A., Edmunds, W. J., Jombart, T., O'Reilly, K., Endo, A., Hellewell, J., Nightingale, E. S., Quilty, B. J., Jarvis, C. I., Russell, T. W., Klepac, P., Bosse, N. I., Funk, S., Abbott, S., Medley, G. F., Gibbs, H., Pearson, C. A. B., Flasche, S., Jit, M., Clifford, S., Prem, K., Diamond, C., Emery, J., Deol, A. K., Procter, S. R., van Zandvoort, K., Sun, Y. F., Munday, J. D., Rosello, A., Auzenbergs, M., Knight, G., Houben, R. M. G. J., and Liu, Y. (2020). Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: A modelling study. *The Lancet Public Health*, 5:e375–e385. 61

Dean, C. B., Ugarte, M. D., and Militino, A. F. (2001). Detecting interaction between random region and fixed age effects in disease mapping. *Biometrics*, 57:197–202. 36

Dean, N., Dong, G., Piekut, A., and Pryce, G. (2019). Frontiers in residential segregation: Understanding neighbourhood boundaries and their impacts. *Tijdschrift voor Economische en Sociale Geografie*, 110:271–288. 134

Department of Health & Social Care (2021a). JCVI interim advice: Potential COVID-19 booster vaccine programme winter 2021 to 2022. https://www.gov.uk/government/publications/jcvi-interim-advice-on-a-potential-coronavirus-covid-19-booster-vaccine-programme-for-winter-2021-to-2022/jcvi-interim-advice-potential-covid-19-booster-vaccine-programme-winter-2021-to-2022. Accessed: 2023-02-15. 129

Department of Health & Social Care (2021b). Joint Committee on Vaccination and Immunisation: advice on priority groups for COVID-19 vaccination, 30 December 2020. https://www.gov.uk/government/publications/priority-groups-for-coronavirus-covid-19-vaccination-advice-from-the-jcvi-30-december-2020/joint-committee-on-vaccination-and-immunisation-advice-on-priority-groups-for-covid-19-vaccination-30-december-2020. Accessed: 2022-06-20. 101

Department of Health & Social Care (2023). Independent report - Chapter 8: Non-pharmaceutical interventions. https://www.gov.uk/government/publications/technical-report-on-the-covid-19-pandemic-in-the-uk/chapter-8-non-pharmaceutical-interventions. Accessed: 2024-11-19. 4

Director-General (2020). WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020. Accessed: 2022-03-01. 57

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20:533–534. 57

dos Santos Siqueira, C. A., de Freitas, Y. N. L., de Camargo Cancela, M., Carvalho, M., Oliveras-Fabregas, A., and de Souza, D. L. B. (2020). The effect of lockdown on the outcomes of COVID-19 in Spain: An ecological study. *PLOS ONE*, 15:e0236779. 60

Duggimpudi, M. B., Abbady, S., Chen, J., and Raghavan, V. V. (2019). Spatio-temporal outlier detection algorithms based on computing behavioral outlierness factor. *Data & Knowledge Engineering*, 122:1–24. 137

Duval, D., Evans, B., Sanders, A., Hill, J., Simbo, A., Kavoi, T., Lyell, I., Simmons, Z., Qureshi, M., Pearce-Smith, N., Arevalo, C. R., Beck, C. R., Bindra, R., and Oliver, I. (2024). Non-pharmaceutical interventions to reduce COVID-19 transmission in the UK: A rapid mapping review and interactive evidence gap map. *Journal of Public Health*, 46:e279–e293. 4

Dziedzic, A., Issa, J., Hussain, S., Tanasiewicz, M., Wojtyczka, R., Kubina, R., Konwinska, M. D., and Riad, A. (2022). COVID-19 vaccine booster hesitancy (VBH) of healthcare professionals and students in Poland: Cross-sectional survey-based study. *Frontiers in Public Health*, 10. 99, 130, 131

Eberly, L. E. and Carlin, B. P. (2000). Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Statistics in Medicine*, 19(17-18):2279–2294. 35

Everitt, B. S., Landau, S., Leese, M., and Stahl, D. (2011). *Cluster Analysis*. Wiley. 55

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, 154:85–100. 115

Frühwirth-Schnatter, S. and Wagner, H. (2011). Bayesian variable selection for random intercept modeling of Gaussian and non-Gaussian data. In: Bayesian statistics 9 (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M.West, eds.). *Oxford University Press*, pages 165–200. 115

Gaffney, A., Himmelstein, D. U., McCormick, D., and Woolhandler, S. (2022). Disparities in COVID-19 vaccine booster uptake in the USA: December 2021–February 2022. *Journal of General Internal Medicine*, 37:2918–2921. 99, 129

Gao, H. and Bradley, J. R. (2019). Bayesian analysis of areal data with unknown adjacencies using the stochastic edge mixed effects model. *Spatial Statistics*, 31:100357. 134

Geers, A. L., Clemens, K. S., Colagiuri, B., Jason, E., Colloca, L., Webster, R., Vase, L., Seig, M., and Faasse, K. (2022). Do side effects to the primary COVID-19 vaccine reduce intentions for a COVID-19 vaccine booster? *Annals of Behavioral Medicine*, 56:761–768. 100

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis (3rd ed.)*. Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC, 2000 Corporate Blvd N.W. Boca Raton, FL 33431, USA. ISBN 978-0-4291-1307-9. 12, 14, 17, 22, 27, 28

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6:721–741. 19

Gerli, A. G., Centanni, S., Miozzo, M. R., Virchow, J. C., Sotgiu, G., Canonica, G. W., and Soriano, J. B. (2020). COVID-19 mortality rates in the European Union, Switzerland, and the UK: Effect of timeliness, lockdown rigidity, and population density. *Minerva Medica*, 111. 59

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Eds., Bayesian Statistics. *Clarendon Press, Oxford*, 4:169–193. 21, 22

Giordani, P., Ferraro, M. B., and Martella, F. (2020). *An introduction to clustering with R*. Springer, Singapore, 1st edition. ISBN 978-981-13-0552-8. 44, 54

Gleditsch, K. S. and Ward, M. D. (2001). Measuring space: A minimum-distance database and applications to international studies. *Journal of Peace Research*, 38:739–758. 34

Gómez-Rubio, V. (2020). *Bayesian inference with INLA*. Chapman & Hall/CRC Press. 117

Goolsbee, A. and Syverson, C. (2021). Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020. *Journal of Public Economics*, 193:104311. 58

Gray, W. K., Navaratnam, A. V., Day, J., Wendon, J., and Briggs, T. W. R. (2021). COVID-19 hospital activity and in-hospital mortality during the first and second waves of the pandemic in England: An observational study. *Thorax*. 94

Grint, D. J., Wing, K., Houlihan, C., Gibbs, H. P., Evans, S. J. W., Williamson, E., McDonald, H. I., Bhaskaran, K., Evans, D., Walker, A. J., Hickman, G., Nightingale, E., Schultze, A., Rentsch, C. T., Bates, C., Cockburn, J., Curtis, H. J., Morton, C. E., Bacon, S., Davy, S., Wong, A. Y. S., Mehrkar, A., Tomlinson, L., Douglas, I. J., Mathur, R., MacKenna, B., Ingelsby, P., Croker, R., Parry, J., Hester, F., Harper, S., DeVito, N. J., Hulme, W., Tazare, J., Smeeth, L., Goldacre, B., and Eggo, R. M. (2021). Severity of severe acute respiratory system coronavirus 2 (SARS-CoV-2) Alpha variant (B.1.1.7) in England. *Clinical Infectious Diseases*. 91, 93

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, 11:1. 135

Hamidi, S., Sabouri, S., and Ewing, R. (2020). Does density aggravate the COVID-19 pandemic? *Journal of the American Planning Association*, 86:495–509. 93

Hartigan, J. A. (1975). *Clustering algorithms*. John Wiley & Sons, Inc., USA, 99th edition. ISBN 978-0-471-35645-5. 47

Hartigan, J. A. and Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, 28:100. 47

Harvey, A. C. (1993). *Time Series Models. 2nd Edition*. MIT Press, Cambridge. 32

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York, second edition. 45, 46, 48, 51, 55

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97. 19

Haynes, W. (2013). Bonferroni correction. *Encyclopedia of Systems Biology*, pages 154–154. 69

Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58(5):295–300. 191

Huang, L., Pickle, L. W., and Das, B. (2008). Evaluating spatial methods for investigating global clustering and cluster detection of cancer cases. *Statistics in Medicine*, 27:5111–5142. 136

Johns Hopkins University & Medicine (2022). COVID-19 dashboard. https://coronavirus.jhu.edu/map.html. Accessed: 2022-06-15. 96, 97

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., and Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, 451:990–993. 56

Karesh, W. B., Cook, R. A., Bennett, E. L., and Newcomb, J. (2005). Wildlife trade and global disease emergence. *Emerging Infectious Diseases*, 11:1000–1002. 57

Kaufman, L. and Rousseeuw, P. J. (1987). Clustering by means of medoids. In *Proceedings of the statistical data analysis based on the L1 norm conference, Neuchatel, Switzerland*, volume 31. 48

Khalili, M., Karamouzian, M., Nasiri, N., Javadi, S., Mirzazadeh, A., and Sharifi, H. (2020). Epidemiological characteristics of COVID-19: A systematic review and meta-analysis. *Epidemiology and Infection*, 148:e130. 92

Khubchandani, J., Sharma, S., Price, J. H., Wiblishauser, M. J., Sharma, M., and Webb, F. J. (2021). COVID-19 vaccination hesitancy in the United States: A rapid national assessment. *Journal of Community Health*, 46:270–277. 99

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in medicine*, 19 17-18:2555–67. 39

Kou, Y., Lu, C.-T., and Chen, D. (2006). Spatial weighted outlier detection. *Proceedings of the 2006 SIAM International Conference on Data Mining*, pages 614–618. 137

Kourlaba, G., Kourkouni, E., Maistreli, S., Tsopela, C.-G., Molocha, N.-M., Triantafyllou, C., Koniordou, M., Kopsidas, I., Chorianopoulou, E., Maroudi-Manta, S., Filippou, D., and Zaoutis, T. E. (2021). Willingness of Greek general population to get a COVID-19 vaccine. *Global Health Research and Policy*, 6:3. 98

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86. 15, 210

Kumari, A., Ranjan, P., Chopra, S., Kaur, D., Kaur, T., Upadhyay, A. D., Isaac, J. A., Kasiraj, R., Prakash, B., Kumar, P., Dwivedi, S. N., and Vikram, N. K. (2021). Knowledge, barriers and facilitators regarding COVID-19 vaccine and vaccination programme among the general population: A cross-sectional survey from one thousand two hundred and forty-nine participants. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 15:987–992. 98, 99, 130

Lawson, A. B. (2018). *Bayesian Disease Mapping*. Chapman and Hall/CRC. 8

Lawson, A. B., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.-F., and Bertollini, R. (1999). *Disease mapping and risk assessment for public health*. John Wiley & Sons Ltd. 7

Lazarus, J. V., Ratzan, S. C., Palayew, A., Gostin, L. O., Larson, H. J., Rabin, K., Kimball, S., and El-Mohandes, A. (2021). A global survey of potential acceptance of a COVID-19 vaccine. *Nature Medicine*, 27:225–228. 129

Ledford, H., Cyranoski, D., and Noorden, R. V. (2020). The UK has approved a COVID vaccine — here's what scientists now want to know. *Nature*, 588:205–206. 97

Lee, D. (2011). A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spatial and Spatio-temporal Epidemiology*, 2:79–89. 33

Lee, D. (2020). A tutorial on spatio-temporal disease risk modelling in R using Markov chain Monte Carlo simulation and the CARBayesST package. *Spatial and Spatio-temporal Epidemiology*, 34:100353. 69, 141

Lee, D. and Mitchell, R. (2013). Locally adaptive spatial smoothing using conditional auto-regressive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62:593–608. 131, 134

Lee, D., Robertson, C., McRae, C., and Baker, J. (2022). Quantifying the impact of air pollution on COVID-19 hospitalisation and death rates in Scotland. *Spatial and Spatio-temporal Epidemiology*, 42:100523. 57

Lee, D., Rushworth, A., and Napier, G. (2018). Spatio-temporal areal unit modeling in R with conditional autoregressive priors using the CARBayesST Package. *Journal of Statistical Software*, 84. 20, 22, 44

Lee, D., Rushworth, A., Napier, G., and Pettersson, W. (2021). CARBayesST version 3.2: Spatio-temporal areal unit modelling in R with conditional autoregressive priors. https://cran.r-project.org/web/packages/CARBayesST/vignettes/CARBayesST.pdf. Accessed: 2022-03-11. 74

Lee, J. and Huang, Y. (2022). COVID-19 vaccine hesitancy: The role of socioeconomic factors and spatial effects. *Vaccines*, 10:352. 99, 130

Leroux, B. G., Lei, X., and Breslow, N. (2000). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In: Halloran M.E., Berry D. (eds) Statistical models in epidemiology, the environment, and clinical trials. *The IMA Volumes in Mathematics and its Applications, vol. 116. Springer, New York, NY*, pages 179–191. 36, 215

Li, M., Shi, X., Li, X., Ma, W., He, J., and Liu, T. (2019). Sensitivity of disease cluster detection to spatial scales: an analysis with the spatial scan statistic method. *International Journal of Geographical Information Science*, 33:2125–2152. 136

Liao, T. F., Bolano, D., Brzinsky-Fay, C., Cornwell, B., Fasang, A. E., Helske, S., Piccarreta, R., Raab, M., Ritschard, G., Struffolino, E., and Studer, M. (2022). Sequence analysis: Its past, present, and future. *Social Science Research*, 107:102772. 102

Liu, R. and Li, G. M. (2021). Hesitancy in the time of coronavirus: Temporal, spatial, and sociodemographic variations in COVID-19 vaccine hesitancy. *SSM - Population Health*, 15:100896. 99

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65:297–303. 32

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28:129–137. 47

Lu, C.-T., Chen, D., and Kou, Y. (2003). Algorithms for spatial outlier detection. In *Third IEEE International Conference on Data Mining*, pages 597–600. IEEE Comput. Soc. 137

Lu, H. and Carlin, B. P. (2005). Bayesian areal wombling for geographical boundary analysis. *Geographical Analysis*, 37:265–285. 134

MacDonald, N. E. (2015). Vaccine hesitancy: Definition, scope and determinants. *Vaccine*, 33:4161–4164. 5, 98

MacNab, Y. C. (2022). Bayesian disease mapping: Past, present, and future. *Spatial Statistics*, 50:100593. 2, 3, 8, 33, 134

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2023). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.6 — For new features, see the 'NEWS' and the 'Changelog' file in the package source). 50

Manley, D. (2021). *Scale, Aggregation, and the Modifiable Areal Unit Problem*, pages 1711–1725. Springer Berlin Heidelberg. 192

Mayer, B., Helm, S., Heinz, E., Barnett, M., and Arora, M. (2022). Doubt in store: Vaccine hesitancy among grocery workers during the COVID-19 pandemic. *Journal of Behavioral Medicine*. 99

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models (2nd ed.)*. Routledge. 11

Mellone, A., Gong, Z., and Scarciotti, G. (2021). Modelling, prediction and design of COVID-19 lockdowns by stringency and duration. *Scientific Reports*, 11:15708. 93

Merriam-Webster Online Dictionary (2022). Lockdown definition and meaning. https://www.merriam-webster.com/dictionary/lockdown. Accessed: 2022-04-14. 58

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21:1087–1092. 19

Mewhirter, J., Sagir, M., and Sanders, R. (2022). Towards a predictive model of COVID-19 vaccine hesitancy among American adults. *Vaccine*, 40:1783–1789. 99, 128

Ministry of Housing, C. . L. G. (2019a). National statistics: English indices of deprivation 2019 (file 1). https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019. Accessed: 2022-03-16. 84

Ministry of Housing, C. . L. G. (2019b). National statistics: English indices of deprivation 2019 (file 10). https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019. Accessed: 2022-02-10. 84

Ministry of Housing, C. . L. G. (2019c). The English indices of deprivation 2019 (IoD2019). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/833959/IoD2019_Infographic.pdf. Accessed: 2022-03-16. 84

Ministry of Housing, Communities and Local Government (2023). Local government structure and elections. https://www.gov.uk/guidance/local-government-structure-and-elections. Accessed: 2024-11-22. 192

Monagin, C., Paccha, B., Liang, N., Trufan, S., Zhou, H., Wu, D., Schneider, B. S., Chmura, A., Epstein, J., Daszak, P., Ke, C., and Rabinowitz, P. M. (2018). Serologic and behavioral risk survey of workers with wildlife contact in China. *PLOS ONE*, 13:e0194647. 56

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley-Blackwell, Hoboken, NJ, 5 edition. ISBN 978-0-4705-4281-1. 11

Moraga, P. (2019). *Geospatial health data: Modeling and visualization with R-INLA and Shiny*. Chapman & Hall/CRC Biostatistics Series. 117

Moraga, P. (2023). *Spatial Statistics for Data Science: Theory and Practice with R*. Chapman & Hall/CRC Data Science Series. 137

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37:17–23. 30, 137

Morens, D. M., Breman, J. G., Calisher, C. H., Doherty, P. C., Hahn, B. H., Keusch, G. T., Kramer, L. D., LeDuc, J. W., Monath, T. P., and Taubenberger, J. K. (2020). The origin of COVID-19 and why it matters. *The American Journal of Tropical Medicine and Hygiene*, 103:955–959. 56

Muegge, R., Dean, N., Jack, E., and Lee, D. (2023). National lockdowns in England: The same restrictions for all, but do the impacts on COVID-19 mortality risks vary geographically? *Spatial and Spatio-temporal Epidemiology*, 44:100559. i

Muegge, R., Jack, E., Dean, N., and Lee, D. (2024). COVID-19 vaccine fatigue in Scotland: How do the trends in attrition rates for the second and third doses differ by age, sex, and council area? *Journal of the Royal Statistical Society Series A: Statistics in Society*. i

Murray, A. T., Grubesic, T. H., and Wei, R. (2014). Spatially significant cluster detection. *Spatial Statistics*, 10:103–116. 136

Nagy, A. and Alhatlani, B. (2021). An overview of current COVID-19 vaccine platforms. *Computational and Structural Biotechnology Journal*, 19:2508–2517. 97

National Portrait Gallery (2024). A picture of health. https://www.npg.org.uk/learning/a-picture-of-health/timeline/. Accessed: 2024-11-20. 1

Naylor, J., Ho, A., and Soiza, R. L. (2024). Why the UK is vaccinating its older adult population against RSV—what geriatricians should know. *Age and Ageing*, 53. 192

Ndwandwe, D. and Wiysonge, C. S. (2021). COVID-19 vaccines. *Current Opinion in Immunology*, 71:111–116. 97

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135:370. 11

NHS (2021). COVID-19 vaccination statistics. https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2021/12/COVID-19-weekly-announced-vaccinations-23-December-2021.pdf. Accessed: 2024-11-19. 5

NHS (2022). How to get a booster dose of the coronavirus (COVID-19) vaccine. https://www.nhs.uk/conditions/coronavirus-covid-19/coronavirus-vaccination/how-to-get-a-coronavirus-vaccine/how-to-get-a-booster-dose/. Accessed: 2022-06-20. 98

NHS (2024). What is public health? https://www.healthcareers.nhs.uk/working-health/working-public-health/what-public-health. Accessed: 2024-11-18. 1

Noordzij, M., Dekker, F. W., Zoccali, C., and Jager, K. J. (2010). Measures of disease frequency: Prevalence and incidence. *Nephron Clinical Practice*, 115:c17–c20. 7

NRS (2021). Mid-2020 population estimates Scotland. https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/popul

ation/population-estimates/mid-year-population-estimates/mid-2020. Accessed: 2022-07-29. 102

ONS (2021a). Coronavirus (COVID-19) infection survey technical article: Waves and lags of COVID-19 in England, June 2021. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsurveytechnicalarticle/wavesandlagsofcovid19inenglandjune2021. Accessed: 2022-03-11. 78, 92

ONS (2021b). Dataset: Death registrations and occurrences by local authority and health board. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/datasets/deathregistrationsandoccurrencesbylocalauthorityandhealthboard. Accessed: 2021-06-02. 65

ONS (2021c). Dataset: Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland. Accessed: 2021-12-01. 66, 101

Openshaw, S. (1983). *The Modifiable Areal Unit Problem*. Concepts and techniques in modern geography. Geo Books. 192

Page, M. L. and McNamara, A. (2021). Alpha COVID-19 variant (B.1.1.7). https://www.newscientist.com/definition/uk-covid-19-variant-b-1-1-7/. Accessed: 2022-03-11. 82

Palladino, R., Bollon, J., Ragazzoni, L., and Barone-Adesi, F. (2020). Excess deaths and hospital admissions for COVID-19 due to a late implementation of the lockdown in Italy. *International Journal of Environmental Research and Public Health*, 17. 59, 60

Palladino, R., Bollon, J., Ragazzoni, L., and Barone-Adesi, F. (2021). Effect of implementation of the lockdown on the number of COVID-19 deaths in four European countries. *Disaster Medicine and Public Health Preparedness*, 15:e40–e42. 59

Paul, E. and Fancourt, D. (2022). Predictors of uncertainty and unwillingness to receive the COVID-19 booster vaccine: An observational study of 22,139 fully vaccinated adults in the UK. *The Lancet Regional Health - Europe*, 14:100317. 100, 130

Pearson, K. (1896). VII. mathematical contributions to the theory of evolution.—III. regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 187:253–318. 29

Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51:406–413. 35

PHS (2022). The vaccines used to protect against coronavirus. https://www.nhsinform.scot/covid-19-vaccine/the-vaccines/the-vaccines-used-to-protect-against-coronavirus/. Accessed: 2022-05-26. 97, 102

Plummer, M., Best, N., Cowles, K., and Vines, K. (2005). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6. 22

Prati, G. and Mancini, A. D. (2021). The psychological impact of COVID-19 pandemic lockdowns: A review and meta-analysis of longitudinal studies and natural experiments. *Psychological Medicine*, 51:201–211. 58

Prime Minister's Office (2021a). Oral statement - PM statement to the House of Commons on roadmap for easing lockdown restrictions in England: 22 February 2021. https://www.gov.uk/government/speeches/pm-statement-to-the-house-of-commons-on-roadmap-for-easing-lockdown-restrictions-in-england-22-february-2021. Accessed: 2022-03-10. 93

Prime Minister's Office (2021b). Prime Minister announces national lockdown. https://www.gov.uk/government/news/prime-minister-announces-national-lockdown. Accessed: 2022-03-10. 93

Ramaswamy, S., Rastogi, R., and Shim, K. (2000). Efficient algorithms for mining outliers from large data sets. *ACM SIGMOD Record*, 29:427–438. 145

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66:846. 53, 54

Rashedi, J., Poor, B. M., Asgharzadeh, V., Pourostadi, M., Kafil, H. S., Vegari, A., Tayebi-Khosroshahi, H., and Asgharzadeh, M. (2020). Risk factors for COVID-19. *Le infezioni in medicina*, 28:469–474. 57

Remuzzi, A. and Remuzzi, G. (2020). COVID-19 and Italy: What next? *The Lancet*, 395:1225–1228. 57

Richardson, S., Thomson, A., Best, N., and Elliott, P. (2004). Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives*, 112:1016–1025. 134

Riebler, A., Sørbye, S. H., Simpson, D., and Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25:1145–1165. 15, 38, 114, 115, 116, 117, 149, 150, 151, 210

Ripley, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press. 30

Robert, C. and Casella, G. (2011). A short history of Markov chain Monte Carlo: Subjective recollections from incomplete data. *Statistical Science*, 26. 17

Rothman, K. (2012). *Epidemiology: An Introduction*. OUP USA. 2

Rothman, K., Greenland, S., and Lash, T. (2008). *Modern Epidemiology*. Wolters Kluwer Health/Lippincott Williams & Wilkins. 191

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65. 50

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields*. Chapman and Hall/CRC. 23

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71:319–392. 23

Rushworth, A., Lee, D., and Mitchell, R. (2014). A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in greater London. *Spatial and Spatio-temporal Epidemiology*, 10:29–38. 42, 44, 62, 71, 72, 75

Rushworth, A., Lee, D., and Sarran, C. (2017). An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 66:141–157. 134

Rzymski, P., Poniedziałek, B., and Fal, A. (2021). Willingness to receive the booster COVID-19 vaccine dose in Poland. *Vaccines*, 9:1286. 99, 128, 129, 130

Sartorius, B., Lawson, A. B., and Pullan, R. L. (2021). Modelling and predicting the spatio-temporal spread of COVID-19, associated deaths and impact of key risk factors in England. *Scientific Reports*, 11:5378. 61, 93

Schubert, E., Zimek, A., and Kriegel, H.-P. (2014a). Generalized outlier detection with flexible kernel density estimates. *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 542–550. 137, 138, 143, 183

Schubert, E., Zimek, A., and Kriegel, H.-P. (2014b). Local outlier detection reconsidered: A generalized view on locality with applications to spatial, video, and network outlier detection. *Data Mining and Knowledge Discovery*, 28:190–237. 137

Scottish Government (2020a). First COVID-19 vaccinations in Scotland take place. https://www.gov.scot/news/first-covid-19-vaccinations-in-scotland-take-place/. Accessed: 2023-02-15. 129

Scottish Government (2020b). Scottish index of multiple deprivation 2020: Introduction. https://www.gov.scot/publications/scottish-index-multiple-deprivation-2020/pages/5/. Accessed: 2023-02-14. 130

Scottish Government (2020c). Scottish index of multiple deprivation 2020v2 - indicators. https://www.gov.scot/publications/scottish-index-of-multiple-deprivation-2020v2-indicator-data/. Accessed: 2024-01-18. 218

Scottish Government (2021a). COVID-19 vaccine deployment plan 2021. https://www.gov.scot/binaries/content/documents/govscot/publications/strategy-plan/2020/12/coronavirus-covid-19-vaccine-deployment-plan-2021/documents/coronavirus-covid-19-vaccine-deployment-plan-2021/coronavirus-covid-19-vaccine-deployment-plan-2021/govscot%3Adocument/COVID-19%2Bvaccine%2Bdeployment%2Bplan%2B14%2BJanuary%2B2021.pdf. Accessed: 2023-02-07. 128

Scottish Government (2021b). Steps to accelerate booster programme. https://www.gov.scot/news/steps-to-accelerate-booster-programme/. Accessed: 2022-07-28. 102

Scottish Government (2022). How many doses of the coronavirus (COVID-19) vaccine am I eligible for. https://www.nhsinform.scot/covid-19-vaccine/how-many-doses-of-the-coronavirus-covid-19-vaccine-am-i-eligible-for. Accessed: 2023-02-16. 130

Sedaghat, L., Hersey, J., and McGuire, M. P. (2013). Detecting spatio-temporal outliers in crowdsourced bathymetry data. *Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*, pages 55–62. 137

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology*, 46:561–584. 69

Shekhar, S. (2003). A unified approach to detecting spatial outliers. *GeoInformatica*, 7:139–166. 137

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., and Siddique, R. (2020). COVID-19 infection: Emergence, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research*, 24:91–98. 56

Silva, L., Filho, D. F., and Fernandes, A. (2020). The effect of lockdown on the COVID-19 epidemic in Brazil: Evidence from an interrupted time series design. *Cadernos de Saúde Pública*, 36. 60

Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32. 15, 16, 17, 37, 111, 115, 117, 149, 210

Singh, A. K. and Lalitha, S. (2018). A novel spatial outlier detection technique. *Communications in Statistics - Theory and Methods*, 47:247–257. 137

Small, H. (2017). *A brief history of Florence Nightingale*. Robinson. 2

Sønderskov, K. M., Vistisen, H. T., Dinesen, P. T., and Østergaard, S. D. (2021). COVID-19 booster vaccine willingness. *Danish medical journal*, 69. 99, 128

Soares, P., Rocha, J. V., Moniz, M., Gama, A., Laires, P. A., Pedro, A. R., Dias, S., Leite, A., and Nunes, C. (2021). Factors associated with COVID-19 vaccine hesitancy. *Vaccines*, 9:300. 128

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64:583–639. 27

Sørbye, S. H. and Rue, H. (2014). Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics*, 8:39–51. 38, 114, 117

Statista (2020). Largest urban agglomerations in the United Kingdom in 2020. https://www.statista.com/statistics/294645/population-of-selected-cities-in-united-kingdom-uk/. Accessed: 2022-03-16. 90

Sudre, C. H., Murray, B., Varsavsky, T., Graham, M. S., Penfold, R. S., Bowyer, R. C., Pujol, J. C., Klaser, K., Antonelli, M., Canas, L. S., Molteni, E., Modat, M., Cardoso, M. J., May, A., Ganesh, S., Davies, R., Nguyen, L. H., Drew, D. A., Astley, C. M., Joshi, A. D., Merino, J., Tsereteli, N., Fall, T., Gomez, M. F., Duncan, E. L., Menni, C., Williams, F. M. K., Franks, P. W., Chan, A. T., Wolf,

J., Ourselin, S., Spector, T., and Steves, C. J. (2021). Attributes and predictors of long COVID. *Nature Medicine*, 27:626–631. 57

Susser, M. and Stein, Z. (2009). *The British Sanitary Movement: Edwin Chadwick*, pages 50–64. Oxford University Press. 2

Swift, L., Hunter, P. R., Lees, A. C., and Bell, D. J. (2007). Wildlife trade and the emergence of infectious diseases. *EcoHealth*, 4:25. 57

Swinney, J. (2021). Health protection (coronavirus) (requirements) (Scotland) amendment (no. 2) regulations 2021. https://www.legislation.gov.uk/ssi/2021/349/pdfs/ssi_20210349_en.pdf. Accessed: 2023-02-13. 129

Takalo, R., Hytti, H., and Ihalainen, H. (2005). Tutorial on univariate autoregressive spectral analysis. *Journal of Clinical Monitoring and Computing*, 19:401–410. 22

Tang, B. and He, H. (2017). A local density-based approach for outlier detection. *Neurocomputing*, 241:171–180. 137, 138, 143, 182, 183

Taylor, L. H., Latham, S. M., and Woolhouse, M. E. (2001). Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356:983–989. 56

Tessema, Z. T., Tesema, G. A., Ahern, S., and Earnest, A. (2023). A systematic review of areal units and adjacency used in Bayesian spatial and spatio-temporal conditional autoregressive models in health research. *International Journal of Environmental Research and Public Health*, 20:6277. 3

The Scottish Parliament (2022). Coronavirus (COVID-19): Vaccinations in Scotland – latest data. https://spice-spotlight.scot/2022/07/27/coronavirus-covid-19-vaccinations-in-scotland-latest-data/. Accessed: 2022-07-29. 97

Thorisdottir, I. E., Asgeirsdottir, B. B., Kristjansson, A. L., Valdimarsdottir, H. B., Tolgyes, E. M. J., Sigfusson, J., Allegrante, J. P., Sigfusdottir, I. D., and Halldorsdottir, T. (2021). Depressive symptoms, mental wellbeing, and substance use among adolescents before and during the COVID-19 pandemic in Iceland: A longitudinal, population-based study. *The Lancet Psychiatry*, 8:663–672. 58

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, 18:267–276. 51, 146

Tinson, A. (2021). What geographic inequalities in COVID-19 mortality rates and health can tell us about levelling up. https://www.health.org.uk/news-and-comment/charts-and-infographics/what-geographic-inequalities-in-covid-19-mortality-rates-can-tell-us-about-levelling-up. Accessed: 2022-03-10. 84

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46:234. 5, 28, 134

Troiano, G. and Nardi, A. (2021). Vaccine hesitancy in the era of COVID-19. *Public Health*, 194:245–251. 99, 129

Ugarte, M., Etxeberria, J., Goicoa, T., and Ardanaz, E. (2012). Gender-specific spatio-temporal patterns of colorectal cancer incidence in Navarre, Spain (1990–2005). *Cancer Epidemiology*, 36:254–262. 41

UK Health Security Agency (2023a). Coronavirus (COVID-19) in the UK. https://coronavirus.data.gov.uk/details/vaccinations. Accessed: 2023-01-04. 97

UK Health Security Agency (2023b). Coronavirus (COVID-19) in the UK: Vaccinations in Scotland. https://coronavirus.data.gov.uk/details/vaccinations?areaType=nation&areaName=Scotland. Accessed: 2023-02-13. 129

United Nations - Department of Economic and Social Affairs (2015). Transforming our world: The 2030 agenda for sustainable development. https://sdgs.un.org/2030agenda. Accessed: 2024-11-20. 1

Usher, K., Durkin, J., and Bhullar, N. (2020). The COVID-19 pandemic and mental health impacts. *International Journal of Mental Health Nursing*, 29:315–318. 57

van Niekerk, J., Krainski, E., Rustand, D., and Rue, H. (2023). A new avenue for Bayesian inference with INLA. *Computational Statistics & Data Analysis*, 181:107692. 23, 24, 25, 26

van Niekerk, J. and Rue, H. (2024). Low-rank variational Bayes correction to the Laplace method. *Journal of Machine Learning Research*, 25(62):1–25. 25, 26

Vandenbroucke, J. P., Broadbent, A., and Pearce, N. (2016). Causality and causal inference in epidemiology: The need for a pluralistic approach. *International Journal of Epidemiology*, 45:1776–1786. 191

Venkatesan, P. (2024). The UK COVID-19 inquiry and critical care. *The Lancet Respiratory Medicine*, 12:e63–e64. 3

Vinod, H. D. (1969). Integer programming and the theory of grouping. *Journal of the American Statistical Association*, 64:506–519. 49

Wakefield, J. (2008). Ecologic studies revisited. *Annual Review of Public Health*, 29:75–90. 8

Wakefield, J. and Salway, R. (2001). A statistical framework for ecological and aggregate studies. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 164:119–137. 131, 218

Waller, L. and Carlin, B. (2010). *Disease Mapping*, pages 217–243. Chapman & Hall/CRC handbooks of modern statistical methods. 8, 10

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *The Journal of Machine Learning Research*, 11:3571–3594. 28

WHO (2022). COVID-19 vaccine tracker and landscape. https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines. Accessed: 2022-05-27. 97

Williamson, E. J., Walker, A. J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C. E., Curtis, H. J., Mehrkar, A., Evans, D., Inglesby, P., Cockburn, J., McDonald, H. I., MacKenna, B., Tomlinson, L., Douglas, I. J., Rentsch, C. T., Mathur, R., Wong, A. Y. S., Grieve, R., Harrison, D., Forbes, H., Schultze, A., Croker, R., Parry, J., Hester, F., Harper, S., Perera, R., Evans, S. J. W., Smeeth, L., and Goldacre,

B. (2020). Factors associated with COVID-19-related death using OpenSAFELY. *Nature*, 584:430–436. 57, 93

Wolff, D., Nee, S., Hickey, N. S., and Marschollek, M. (2021). Risk factors for COVID-19 severity and fatality: A structured literature review. *Infection*, 49:15–28. 57

Wood, S. N. (2017). *Generalized additive models: An introduction with R (2nd ed.)*. Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC, 2000 Corporate Blvd N.W. Boca Raton, FL 33431, USA. ISBN 978-1-3153-7027-9. 11

World Health Organization (2024). The top 10 causes of death. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. Accessed: 2024-11-19. 3

Yan, B., Zhang, X., Wu, L., Zhu, H., and Chen, B. (2020). Why do countries respond differently to COVID-19? A comparative study of Sweden, China, France, and Japan. *The American Review of Public Administration*, 50:762–769. 94

Zhang, K., Hutter, M., and Jin, H. (2009). A new local distance-based outlier detection approach for scattered real-world data. In *Advances in Knowledge Discovery and Data Mining*, pages 813–822, Berlin, Heidelberg. Springer Berlin Heidelberg. 145