Technische Universität München

TUM School of Life Sciences

Tailored statistical methodology for large-scale registry-based clinical data to optimize online risk prediction

Gregor M. Miller

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereichten Dissertation.

Vorsitz:	Prof. Dr. Percy A. Knolle
Prüfer der Dissertation:	1. Prof. Donna Ankerst, Ph.D.
	2. apl. Prof. Dr. Norbert H. Hüser
	3. Prof. Dr. Ulrich Mansmann

Die Dissertation wurde am 26.08.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 11.01.2023 angenommen.

Abstract

Repeated measurements over time are highly important data for medicine, and for the life sciences in general. The modelling of natural processes, such as the spreading of a virus or the survival of a population, depends on such data. It is essential to estimate the association of risk factors and to predict the development of such processes. However, longitudinal data poses various challenges depending on the outcome of interest. For two fields, survival analysis and count models, we showed how analyses have to be adapted to account for multiple outcomes, to select suitable covariates, and to derive predictions.

In our first field of application, we determined the significance of association between allocation type of kidney grafts and transplant outcomes via survival analysis. Furthermore, we built a publicly available prediction tool estimating outcome risks, thereby facilitating the decision of whether to accept a graft for transplantation or not. Our second field of application was concerned with temporal and spatial information of the COVID-19 pandemic. We derived the methodological approach and explored the capabilities of a fundamental Bayesian methodology to model an autoregressive count process. But we also performed a frequentist cross-sectional analysis of mortality to determine spatially dynamic covariate effects. With a longitudinal analysis we estimated the association of time-dynamic risk factors and we built a model to perform weekly forecasts.

We expect that our transplant risk tool will be an important application for graft allocation decisions and will motivate future external validation and adaptations with respect to new data. In our studies on count models, we extracted a number of crucial pitfalls concerning covariate and model selection, which are superficially not evident and may easily produce misleading results. Findings from such models need to be viewed critically and future autoregressive count analyses should be adapted accordingly.

Zusammenfassung

Wiederholte Messungen im Laufe der Zeit liefern wichtige Daten für die Medizin und die Biowissenschaften im Allgemeinen. Die Modellierung natürlicher Prozesse, wie die Ausbreitung eines Virus oder das Überleben einer Population, basiert auf solchen Daten. Im Hinblick auf solche Prozesse ist es essenziell, Assoziationen von Risikofaktoren abschätzen und Entwicklung vorhersagen zu können. Zeitaufgelöste Daten stellen jedoch je nach Zielvariable verschiedene Herausforderungen dar. Für zwei Bereiche, Überlebensanalysen und Zählmodelle, zeigen wir, wie die Analysen angepasst werden müssen, um mehrere Zielvariablen zu berücksichtigen, geeignete Kovariaten auszuwählen und Vorhersagen abzuleiten.

In unserem ersten Anwendungsbereich haben wir die Signifikanz der Assoziation zwischen Zuweisungsart von Nierentransplantaten und Transplantationsergebnissen mittels Überlebensanalyse ermittelt. Darüber hinaus haben wir ein öffentlich zugängliches Prognosetool entwickelt, das Risiken abschätzt und damit die Entscheidung für oder gegen ein Transplantat erleichtert. Unser zweiter Anwendungsbereich befasste sich mit zeitlichen und räumlichen Informationen der COVID-19-Pandemie. Wir leiteten den methodischen Ansatz ab und untersuchten die Möglichkeiten einer grundlegenden Bayes'schen Methodik zur Modellierung eines autoregressiven Zählprozesses. Wir führten auch eine frequentistische Querschnittsanalyse der Sterblichkeit durch, um räumlich dynamische Kovariateneffekte zu ermitteln. Mit einer Längsschnittanalyse schätzten wir die Assoziation von zeitdynamischen Risikofaktoren und erstellten ein Modell für wöchentliche Prognosen.

Wir erwarten, dass unser Transplantationsrisikoinstrument eine wichtige Anwendung für Entscheidungen über die Zuteilung von Transplantaten sein wird und zukünftige externe Validierungen und Anpassungen an neue Daten motivieren wird. In unseren Analysen zu Zählmodellen haben wir eine Reihe entscheidender Fallstricke bei der Auswahl von Kovariaten und Modellen herausgestellt, die oberflächlich betrachtet nicht offensichtlich sind und leicht zu irreführenden Ergebnissen führen können. Ergebnisse aus solchen Modellen müssen kritisch betrachtet werden und künftige autoregressive Zählanalysen sollten entsprechend angepasst werden.

List of included publications

Peer-reviewed

- Assfalg, V.*, **Miller, G.***, Stocker, F., van Meel, M., Groenevelt, T., Tieken, I., Ankerst, D., Renders, L., Novotny, A., Hartmann, D., Jell, A., Rahmel, A., Wahba, R., Mühlfeld, A., Bouts, A., Ysebaert, D., Globke, B., Jacobs-Tulleneers-Thevissen, D., Piros, L., Stippel, D., Heller, K., Eisenberger, U., van Laecke, S., Weimer, R., Rosenkranz, A. R., Berger, S., Fischer, L., Kliem, V., Vondran, F., Sester, U., Schneeberger, S., Harth, A., Kuypers, D., Függer, R., Arnol, M., Christiaans, M., Weinmann-Menke, J., Krüger, B., Hilbrands, L., Banas, B., Hakenberg, O., Minnee, R., Schwenger, V., Heyne, N., van Zuilen, A., Reindl-Schwaighofer, R., Lopau, K., Hüser, N. & Heemann, U. Kidney Transplantation After Rescue Allocation—the Eurotransplant Experience: A Retrospective Multicenter Outcome Analysis. *Transplantation* **106**, 1215–1226 (2022).

- **Miller, G.**, Ankerst, D. P., Kattan, M. W., Hüser, N., Vogelaar, S., Tieken, I., Heemann, U. & Assfalg, V. Kidney Transplantation Outcome Predictions (KTOP): A Risk Prediction Tool for Kidney Transplants from Brain-dead Deceased Donors Based on a Large European Cohort. *European Urology.* ISSN: 0302-2838 (2022).

*Joint first authors.

In submission

- **Miller, G.**, Menzel, A. & Ankerst, D. P. Association between short-term exposure to air pollution and COVID-19 mortality in all German districts: The importance of confounders. *In submission.* (2022).

Contents

1 Introduction **1**
1.1 Kidney transplantations 2
1.2 COVID-19 4
1.2.1 The Development of COVID-19 4
1.2.2 COVID-19 in Germany 5
1.2.3 Epidemiological metrics of German districts 7

2 Methodological background **13**
2.1 Survival analysis methods 13
2.1.1 The Cox proportional hazards model 14
2.1.2 Competing risks 17
2.1.3 Model diagnostics 19
2.1.4 Clustering in survival analysis 22
2.2 Published COVID-19 time series models 24
2.2.1 Models of early developments in China 24
2.2.2 The IHME model 29
2.2.3 The University of Texas model 31
2.2.4 The Imperial College model 33
2.3 Modelling options for longitudinal count data with autoregression 37
2.3.1 Simple Poisson Model 37
2.3.2 Negative Binomial Model 48
2.3.3 Selected model metrics 52

3 Results **55**
3.1 Kidney transplantation after rescue allocations 56
3.2 A risk prediction tool for kidney transplants 57
3.3 Associations of air pollution and COVID-19 in a cross-sectional analysis 59
3.4 Longitudinal analyses of COVID-19 in Germany 61
3.4.1 Data 62
3.4.2 Methods 63
3.4.3 Results 65

3.4.4 Limitations . . . 68
3.4.5 Conclusion . . . 69

4 Discussion . . . 71

A List of Acronyms . . . 75

B Remarks on Implementation . . . 77
B.1 `tsglm` output . . . 77
B.2 Metropolis sampler code . . . 78
B.3 Stan code . . . 79
B.4 Simulations . . . 81

C Remarks on application . . . 83
C.1 Random covariates in Poisson models . . . 83
C.2 Intervention data . . . 84

List of Figures . . . 87

List of Tables . . . 89

Bibliography . . . 91

Chapter 1

Introduction

Time series and longitudinal data are sequences of measurements of one or multiple variables at equally spaced or varying time intervals. Such repeated observations have a high value in medicine and the life sciences in general due to their ability to capitalize on dependence for improved prediction. Monitoring changes in populations or ongoing natural processes are typical examples of how repeated measures are generated. Daily deaths, hourly averages of wind speed, yearly populations of bark beetles, monthly counts of carcinoma diagnoses are examples of such observations. A frequent objective is to identify and quantify the association of explanatory variables with changes in these processes. However, observations of temporal processes often cannot be analysed independently since previous states influence later ones. Therefore, dependent on the type of data, the application of different approaches is needed to model temporal behaviour and to evaluate explanatory variables [1].

We consider two types of temporal data in this dissertation, patient-level survival and population-level infection counts. We present multiple ways to approach such data and illustrate them in two applications. In the first application, we consider survival and secondary outcomes for patients who received kidney transplants at varying time points during a fixed period. In the second application, we model the infection and death counts of the COVID-19 pandemic.

Specifically, the following sections 1.1 and 1.2 provide an introduction to the two application topics, kidney transplants and spread of COVID-19. The chapters introduce the relevance of the topics, explain respective backgrounds, and thereby show some of the important aspects that might need to be considered in analyses and comparable situations. In chapter 2, we describe tools suitable for survival analysis, including the challenges of competing risks and clustering of individuals due to different outcome events and single individuals receiving or donating multiple grafts. We review leading studies and methodologies investigating COVID-19 temporal dynamics. We develop and evaluate options to model processes on Bayesian foundations and compare their

computational efficiency with frequentist tools. In chapter 3, we summarize our first publication of the thesis investigating the association of risk factors with kidney transplant outcomes based on 27,903 grafts from the large Eurotransplant registry, followed by our second publication developing a publicly available tool to predict patient and graft survival. We present submitted work regarding risk factors concerning COVID-19 mortality across all 400 districts of Germany and show in detail an association analysis of weekly infection counts. Finally, on the basis of these analyses, we discuss the findings, characteristics of our approaches, and possibilities to advance the applied methods in chapter 4.

1.1 Kidney transplantations

Chronic kidney disease is a major contributor to mortality from non-communicable diseases. Nationally representative studies in Australia [2], the USA [3], and Norway [3] have shown that more than 10% of the adult population has a marker for kidney disease. In 2017, about 700 million individuals suffered from the disease, more than from diabetes, asthma, or depressive disorders, and 1.2 million died from chronic kidney disease [4]. The most common causes of chronic kidney disease are hypertension and diabetes. For people with end-stage kidney disease, renal replacement therapy, through dialysis or renal transplantation, is a life-saving but expensive treatment. In 2010, about 2.6 million people worldwide received renal replacement therapy, with estimates suggesting that, at best, only half of all people who needed such a treatment had access to it. The number of people receiving this treatment is expected to more than double to 5.4 million by 2030 [5].

Stable or declining rates of organ donation coupled with increasing end-stage renal disease prevalence have led to a shortage of transplanted kidneys. European data shows that a major barrier to kidney transplantation is low availability of transplants [6]. This is a critical issue because waiting time on dialysis is the strongest modifiable risk factor for renal transplant outcomes [7]. In the USA, patients who receive kidney transplantation are at increased risk following surgery, for example, of new-onset diabetes [8] or cancer at a wide variety of sites [9, 10]. Nonetheless, there is ample evidence that transplants, compared with remaining on dialysis, offer significantly lower mortality, especially when considered over the long term [11–13]. The risk of cardiovascular events was lower and quality of life was substantially better for transplant recipients as shown in an international meta-analysis by Tonelli *et al.* [13]. For these reasons, kidney transplant allocation must be maximally efficient so that organs can be made available

to recipients as quickly as possible while maintaining or even improving the quality of matching between donor and recipient.

Typical outcomes of interest following kidney transplantation are patient and graft survival [14]. A number of risk factors have been well studied and are significantly associated with these outcomes. These include recipient age, recipient gender [15, 16], diabetes, adult polycystic kidney disease [15], frailty [17], depression [18], donor age, immunological matching, ethnic background, and time spent on dialysis [19]. There is evidence that some risk factor effects diminished over time such as, recipient obesity [14] and immunological matching [20]. The role of prolonged cold ischemia time as a risk-increasing effect was not entirely conclusive in the long term [19, 21, 22]. The relevance of many of these risk factors, such as recipient age, recipient gender, diabetes, and immunological matching, was confirmed by studies analysing specifically the Eurotransplant registry [23, 24], and had similar associations to our studies [25, 26].

Depending on these factors, any patient undergoing transplantation is at risk of early death and graft loss. Thus, whenever a transplant becomes available for a particular patient, the treating physician needs to assess the match between graft and recipient and decide whether a transplantation should be performed. Although, many of the risk factors are well known, considering all of them and their interactions to make a reliable decision is a challenge. A few tools have already been developed to facilitate this evaluation by predicting survival chances. However, these often consider only a specific group of risk factors, have a very short prediction period, consider only a certain outcome, or are not transferable to the European context.

Other outcomes after kidney transplantation, which are sometimes studied but not the focus of our analyses, include acute rejection, delayed graft function and primary non-function [14, 15, 27]. Acute rejection can manifest itself in multiple processes, resulting in large heterogeneity in severity and duration [28]. Delayed graft function is usually defined as the use of dialysis within seven days of transplantation [27]. Primary non-function is defined by the failure of the transplanted kidney to function within a certain time post-transplant [29]. Many studies also examine the impact of intermediate outcomes including post-transplant measurements on patient and graft survival. For example, early signs of impaired graft function adversely affect graft survival [15, 16, 27] and elevated serum cholesterol levels at one year after transplant are negatively associated to graft and patient survival [30].

In the critical structure of graft allocation, there were several areas where COVID-19 had a significant impact. Organ procurements declined due to problematic transfer of organs for example due to cancellation of flights, limited or delayed COVID-19 testing capabilities, limited operating room availability, and avoidance of non-essential clinical

activities [31]. Strongly reduced transplant numbers were reported internationally [32], even though a single-centre study in the US of 136 patients suggested that infected patients on the waiting list were more likely to require hospitalization and were at a higher risk of mortality than transplant patients [33]. Kidney transplant waitlist deaths increased by 43% in the US [34], yet COVID-19 severity in kidney transplant recipients was similar to non-transplant patients when accounting for comorbidities suggesting that long-term immunosuppression did not affect the severity of an infection as shown by a French, multi-centre study including 2,878 patients [35]. Aubert *et al.* [36] compared the number of kidney, liver, lung, and heart transplants from 22 countries including Eurotransplant data between 2019 and 2020. Even though the results were heterogeneous between the countries, they found that kidney transplantations showed the largest decrease across nearly all countries. This decrease in transplants lead to an estimated 48,000 patient life-years lost.

1.2 COVID-19

The COVID-19 pandemic underwent a very dynamic development, particularly in the beginning, while many properties were unclear. The following sections will shortly characterize the COVID-19 pandemic to provide enough background to later apply suitable methods on a corresponding application. Firstly, we summarize the spreading of the virus on a global scale. Then, we focus on the development in the first few months in Germany. Finally, we supplement this description by selected quantifications.

1.2.1 The Development of COVID-19

A cluster of 27 pneumonia cases with unknown cause was identified by the end of December, 2019, in Wuhan City, Hubei, a city in central China with a population of over 11 million. Dry cough, shortness of breath, fever, and lung infiltrates were the most notable symptoms displayed. All cases could be linked to a single market, the Huanan Seafood Wholesale Market, trading in fish and other live animal species, including poultry, bats, and snakes [37]. The agent was determined as a novel coronavirus and named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), while the disease was termed COVID-19. It was a zoonotic disease, meaning that the pathogen, in this case the virus, jumped from animal to human [38].

In 2002, another zoonotic coronavirus, SARS-CoV, had already surfaced in the province Guangdong in southern China. Eight years later the Middle East Respiratory Syndrome (MERS-CoV) had emerged. Each of these two coronaviruses had killed

around 800 people with infection numbers in the thousands [38].

However powerful the impression the previous coronaviruses left, their effect was dwarfed by the dramatic impact of SARS-CoV-2. The French Ministry of Health confirmed the first three cases of COVID-19 in Europe on January 24, 2020, only a few weeks after the initial reports from China. Six days later, on January 30, the World Health Organization (WHO) declared the outbreak an International Public Health Emergency. Only weeks later the number of cases surpassed 100,000 infected people in 100 countries, leading the WHO to declare COVID-19 officially a pandemic on March 11 [39]. The number of infections continued to rise more than a hundred-fold, amounting to hundreds of thousands of deaths [40].

Several issues preventing the mitigation of the viral spread of COVID-19 were determined, including the lack of transparency as well as delays in travel restrictions, quarantine, and emergency announcements. Still, millions of dollars were mobilized internationally in response to the virus with respect to multiple fields, including testing ability and capacity, clinical management, and vaccine research. Notably, many regions introduced general measures to minimize social contact and thereby restrain personal lives. For example, many countries restricted travel and established quarantines for arriving travellers and inhabitants of specific areas, affecting 50 million people in China alone [37].

Although early reports and studies struggled with unreliable and sparse data, some properties of the disease have crystallized. Similar to the first reported symptoms, fever, dry cough, shortness of breath, and muscle ache proved to be typical of COVID-19 [41]. Early numbers suggested a mortality rate of 3.4%, a mean incubation period of 5.2 days and a median duration from onset of symptoms to death of 14 days. Older adults over 70 years of age were identified as especially vulnerable to SARS-CoV-2. Months after the initial discovery of the disease there was still no effective antiviral treatment available [37].

1.2.2 COVID-19 in Germany

The first Chinese COVID-19 death was reported on January 11 and eleven days later there were already 470 Chinese infections, such that all train and flight connections to Wuhan were cancelled. The first German infection was reported on January 27 in the Bavarian county of Starnberg. Two weeks later, travellers from the Chinese region Hubei were quarantined and 15 more cases were reported, most of them in Bavaria and many of them connected to an automotive supplier in Starnberg. Meanwhile, the number of COVID-19 deaths internationally had already exceeded those of SARS and

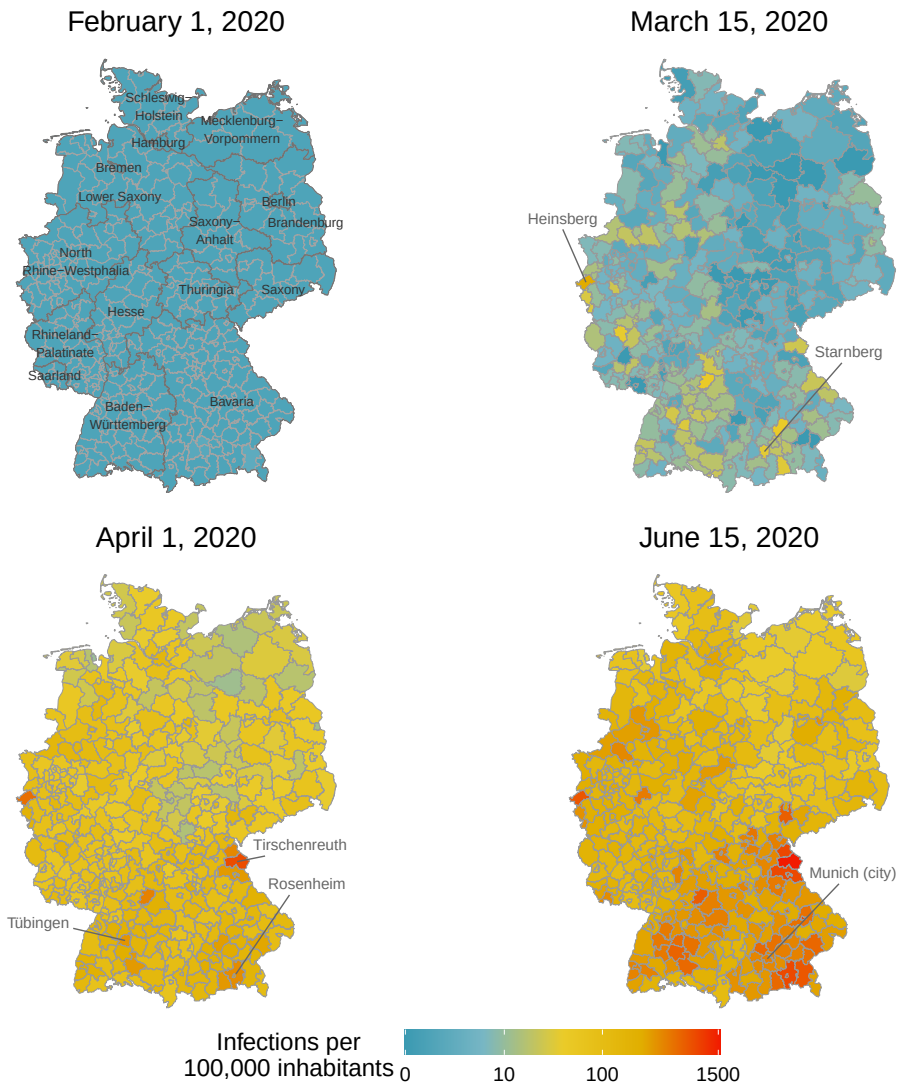

Figure 1.1: Cumulative infections per 100,000 inhabitants in German districts for four different dates. The top left figure shows the German state names, while the others highlight notable districts.

MERS, with more than 40,000 infections in 26 affected countries. By February 26, multiple European countries, including Germany, continued to oppose general closure of country borders. By the end of February, 70 infections were reported all over Germany, including a hotspot with 60 infections in the district Heinsberg in North Rhine-Westphalia at the Western border of Germany [42]. By this time, Italy had the most infections of all European countries, with whole cities completely locked down as well as companies and schools shut down [43, 44].

On March 2, the German Ministry for Health recommended adherence to general hygiene advice. Big events with more than 1,000 participants were advised to be cancelled on March 10. Some states heeded this advice and completely canceled bigger events. At this stage, there were over 500 infections reported. Advice to avoid social contacts and events followed two days later. Starting March 16 many states closed schools and nurseries. Non-essential shops began to shut down and most borders closed. Similar measures were introduced in Austria and France. Four days later several German states restricted the number of people allowed to meet. At the end of March, Germany counted 64,000 infections and 560 deaths [43, 45].

The central institution of the German government for disease monitoring and prevention, the Robert Koch-Institute (RKI), changed their advice to start wearing masks on April 2. On April 6, restrictions on incoming travellers from mid-March were further tightened. Austria began loosening the interventions on April 14. The German government opened the first non-essential shops on April 20 and allowed school attendance starting May 4, however, states implemented individual rules of loosening the measures. In the following days, multiple states decided to make the wearing of masks compulsory, while the other states followed later. At the end of April, a number of public places opened, although the number of infections in Germany had exceeded the threshold of 150,000. Neighbouring countries, such as Austria, had already dropped some of their interventions, and others, such as Italy and France, were planning to relax them in the beginning of May [46].

1.2.3 Epidemiological metrics of German districts

In order to visualize the space-time evolution of the COVID-19 pandemic in Germany we applied the packages `ggplot` [47], `wesanderson` [48], and `ggpubr` [49] to data obtained from the RKI [50] in R [51]; results are shown in Figure 1.1. Although infections were reported in Bavaria starting on January 28, normalized by district, the number of infections remained under 10 per 100,000 inhabitants in the following weeks. By mid-March, circumstances had changed. While some areas remained largely unaffected, the two southern states of Bavaria and Baden-Württemberg, as well as the western border, had a high number of more intensely affected districts, including Starnberg in the south, one of the first infected districts, and Heinsberg in the west, one of the first districts to report a large cluster of infections.

Two weeks later, by April 1, the number of infections rose steeply throughout the country (Figure 1.1). Both infection and death count increased sharply, as shown in Figure 1.2. Initially, deaths lagged behind infections. However, in late April the number

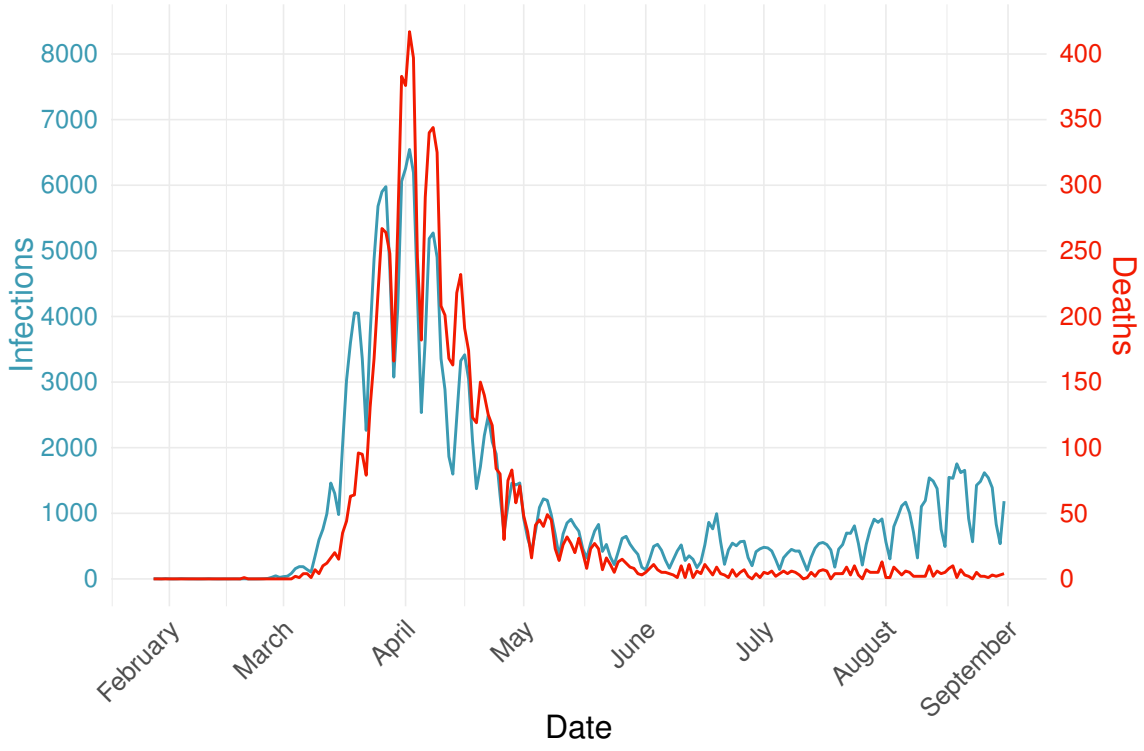

Figure 1.2: Daily infections and deaths in Germany from January 28 to August 31, 2020.

of new cases started to decrease, with deaths at a faster rate that no longer tracked the rate of infections. The reason could be that testing capacity was no longer limited to the more severe cases [52]. The RKI determined that for weeks 12 through 16 of 2020, from mid-March to mid-April, the proportion of positive tests ranged from 6 to 9%. Although the number of weekly performed tests remained similar, at around 300,000 to 400,000, during weeks 22 to 25, spanning the end of May to mid-June, the proportion of positive tests dropped to 0.8-1.3%. These statistics implied that for a steady amount of testing, the proportion of positive outcomes dropped significantly between the end of March and the start of June.

Figure 1.1 shows that after the start of April, all districts had increasing numbers of infections. However, there were some counties in the south and south-east that witnessed particularly elevated infection rates, such as Tübingen, Tirschenreuth, and Rosenheim. These had already comparably high numbers in mid-March, but now their neighbouring districts indicated increases as well.

A characteristic of the COVID-19 metrics in Germany as well as other countries is that reported cases are lower near weekends compared to other weekdays (Figure

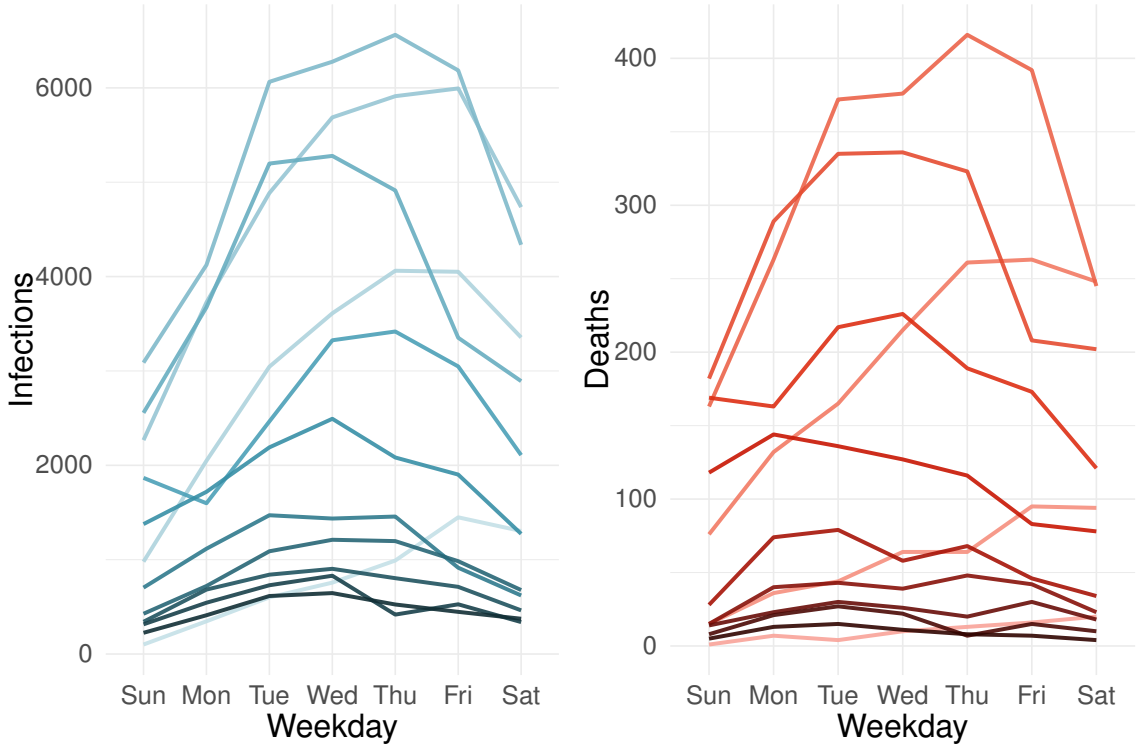

Figure 1.3: Daily reported number of infections (left) and deaths (right) from week of March 3 (light) to May 30 (dark).

1.3). The tendency is not as clear for deaths, but still visible for some of the weeks. Comparable observations have already been made for other diseases, for example, Wei *et al.* [53] identified a weekend-reporting effect on hand, foot, and mouth disease and parotitis infections. Bergman *et al.* [54] attributed periodic weekly oscillation for COVID-19 incidence and mortality in New York and Los Angeles to "biasing practices in case reporting" rather than biological mechanisms.

The distribution of infections in Germany until August 31, 2020, showed systematic differences between age and gender (Figure 1.4). The number of infections was higher for the age groups between 15 and 59 then dropped off slightly for higher age. However, mortality was higher in older age groups and for males compared to females. Summarizing the numbers over the districts to consider the overall case fatality rate, this effect became more apparent. In contrast to mortality, the fatality rate normalizes the number of deaths based on the number of infections. Table 1.1 shows that while the fatality rate was close to zero for young ages, it increased for females and males to 5 and 10%, respectively, for the age group 60 to 79 years. It rose to 22% and 34% for

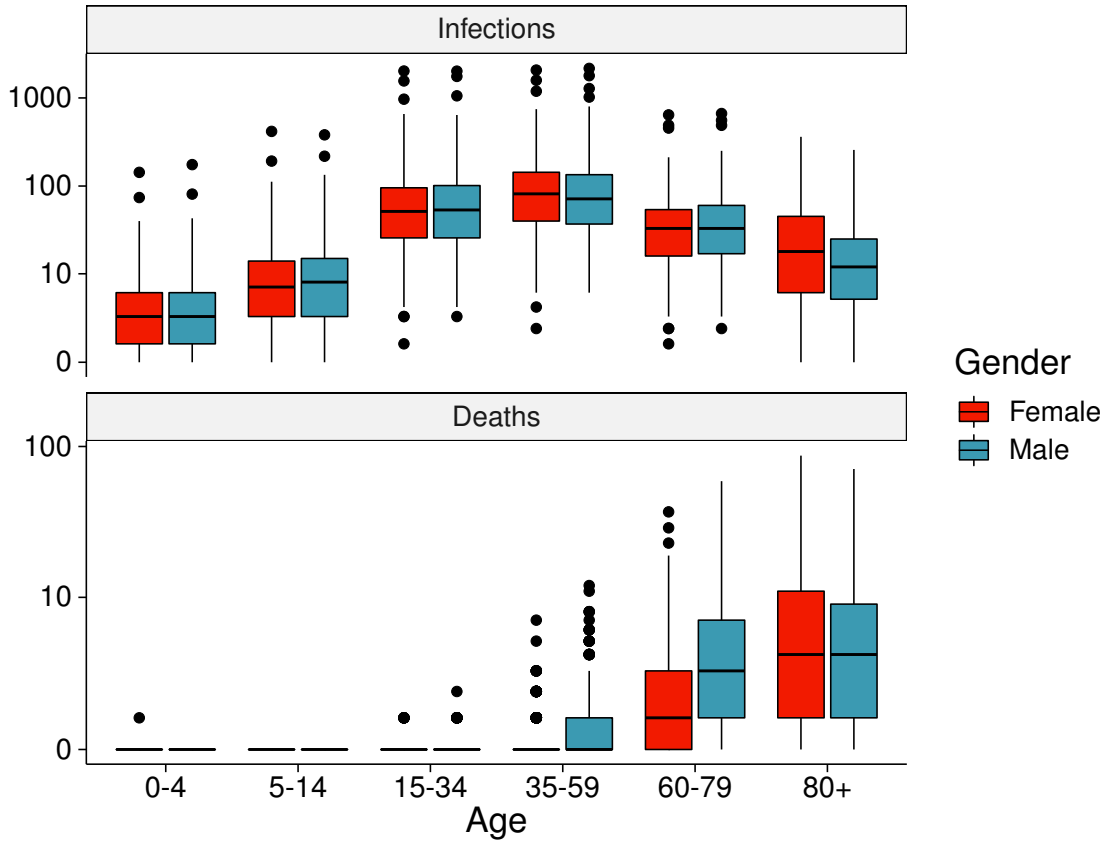

Figure 1.4: Boxplots of reported infections and deaths across the German districts during the period January 28 to August 31, 2020; less than 1% of the respective counts with unknown gender or age are not shown.

80+ year old females and males, respectively. We collapsed the three youngest age groups and applied a Poisson model to deaths with infections as offset and age and gender. Both were significant at the 0.05 level. We note that the number of infections was similar between females and males for all age groups except for the 80+ group with approximately 14,000 female and less than 8,000 male cases. This selection differential might be confounding the observed significant age effect on COVID-19 mortality.

In general, early data from the CDC COVID-19 Response Team [55] in the US showed that the fatality rate increased for older adults ranging from 3% to 11% for 65-84 years followed by 10% to 27% for >84 years. The meta-study of Dorjee & Kim [56] comprising 44 studies and over 20,000 COVID-19 patients identified that next to pre-existing conditions, male gender and age $\geq$ 60 were significantly associated with death and severe disease. Williamson *et al.* [57] considered confounding factors including

	Female			Male		
Age	Infections	Deaths	Case fatality rate [%]	Infections	Deaths	Case fatality rate [%]
0-4	1962	1	0.1	2150	0	0.0
5-14	5024	0	0.0	5373	0	0.0
15-34	34485	5	0.0	36713	13	0.0
35-59	49591	118	0.2	47694	320	0.7
60-79	18356	920	5.0	20211	2085	10.3
80+	14109	3109	22.0	7984	2750	34.4

Table 1.1: COVID-19 metrics by age and gender in Germany until August 31, 2020. Cases with unknown age or gender are not shown.

smoking or obesity in addition to co-morbidities and found that older age and male gender were strongly associated with COVID-19 mortality. Sama *et al.* [58] examined ACE2, a functional receptor for coronaviruses, and found that higher expression levels in men may be the reason for their higher susceptibility and mortality.

Chapter 2

Methodological background

In the following, we describe and develop multiple methods to analyse temporal data, in particular, we lay the groundwork for much of the statistical methodology necessary for the analysis of survival and count data. Specifically, the first section addresses survival analysis and certain challenges associated with it. Later we present various approaches that were used and published to model time series of infectious diseases, especially with respect to COVID-19. Lastly, we focus on Bayesian methods, explore their capacity to model autoregressive count processes, and compare their efficiency with frequentist options.

2.1 Survival analysis methods

Survival analysis is concerned with the time to a particular event and its comparison between specific groups, which are signified by certain characteristics. Without loss of generality, we will refer to survival analysis primarily as the analysis of time to graft loss. But, in general, survival may also refer to other events, for example, patient death, the time to development of a specific disease, or regression of a tumour. Furthermore, survival analysis is not only relevant to medical applications but is also related to analyses in economics, sociology, and engineering, for example, to estimate when a machine will break down. With such a broad field of application and often very particular challenges, the development of very specific tools is necessary.

In the following, we focus largely on the methods we employed in survival analyses in Assfalg *et al.* [25] and Miller *et al.* [26]. We describe some basics of survival analysis and particularly the Cox proportional hazards model as basis of our analyses. Further, we show how this model can be changed and extended to account for challenges including clustering as arises due to organs from the same donor sent to different recipients or recipients receiving multiple organs and multiple competing event outcomes, such as graft loss and death with functioning graft. Additionally, we present an approach to

the selection of suitable covariates as well as to accommodate missing data. For the final models, we discuss how we implemented calibration, decision curve analysis, and accuracy measurements.

2.1.1 The Cox proportional hazards model

The Cox proportional hazards model is one of the most widely used methods in the field of survival analysis. The following description of the basic model is largely based on Therneau & Grambsch [59]. Let X be a matrix of covariates, where x_{ij} is the jth covariate of the ith graft with $i = 1, \ldots, n$, and $j = 1, \ldots, p$. Let x_i be the covariate vector for graft i. The covariates may be fixed or may vary over time. In the latter case, we will write $x_i(t)$. A major goal of survival analysis is to estimate one of its most fundamental functions, the hazard function $h(t)$. It is defined as

$$h(t) = \lim_{\delta \to 0} \frac{P(t < T < t + \delta | T > t)}{\delta}.$$

Thus, the hazard function is the probability that the considered graft fails in the next small interval of time, assuming it survived until time t, divided by the length of this interval δ [60]. For graft i, the Cox model defines the hazard at time t as

$$h_i(t) = h_0(t) \exp(x_i(t)\beta),$$

where h_0 is a non-negative function called the baseline hazard, the hazard of a graft with all covariates equal to zero, and β is a vector of coefficients of length p. In order to estimate β, the partial likelihood was introduced. In the case of untied failure times, where each event time corresponds to exactly one event, the partial likelihood is defined as

$$PL(\beta) = \prod_{i=1}^{n} \prod_{t \geq 0} \left(\frac{y_i(t) r_i(\beta, t)}{\sum_j y_j(t) r_j(\beta, t)} \right)^{dN_i(t)},$$

where $y_i(t)$ is the indicator function displaying if graft i is still under observation at time t and $r_i(\beta, t) = \exp(x_i(t)\beta)$ is the graft-specific risk score. N_i is the number of observed failures at time t, which means it is 0 until failure and 1 thereafter, if observed, and $dN_i(t)$ is the increment in N_i over the infinitesimal time interval $[t, t + dt)$. β can be calculated by maximizing the partial likelihood via the first and second derivative and the application of the Newton-Raphson algorithm. There are a number of methods that can deal with tied failure times, which are a common occurrence in practice. For example, the Efron approximation, which is probably the most prominent method, weights the risk

scores of the affected grafts in the denominator of the partial likelihood by the number of tied grafts for the corresponding time points.

The Cox model is called a proportional hazards model since the following fraction for two grafts with fixed covariates x_i and x_j is constant over time:

$$\frac{h_i(t)}{h_j(t)} = \frac{h_0(t)\exp(x_i\beta)}{h_0(t)\exp(x_j\beta)} = \frac{\exp(x_i\beta)}{\exp(x_j\beta)}.$$

Denote x_i in the following as the ith covariate value for some graft. Assume that we want to test the effect of a treatment in a group A, where $x_1 = 1$, in comparison to a control group B without the treatment, $x_1 = 0$. In our case, this can be, for example, the comparison between a normal and an alternative allocation process of transplant grafts and their association with the subsequent survival. The hazards of grafts in the respective groups, which have otherwise equal covariates, are then:

$$h_A(t) = h_0(t)\exp(\beta_0 + \beta_1 + \beta_2 x_2 + \ldots \beta_p x_p)$$
$$h_B(t) = h_0(t)\exp(\beta_0 + \beta_2 x_2 + \ldots \beta_p x_p).$$

Thus, the hazard ratio of the treatment is

$$\frac{h_A(t)}{h_B(t)} = \exp(\beta_1),$$

for any $t \geq 0$, as shown above. Therefore, the hazard at any time for grafts with the treatment is $\exp(\beta_1)$ times that of grafts without the treatment [61].

In order to better understand the effect on survival, we consider the survival function

$$S(t) = P(T > t) = \exp\left(-\int_0^t h(u)du\right) = \exp(-H(t)), \quad 0 < t < \infty,$$

giving the probability of surviving up to a point t, where $H(t) = \int_0^t h(u)du$ is the cumulative hazard function [60]. It follows for the survival of group A, which has received the

treatment, that [61]

$$\begin{aligned} S_A(t) &= -\int_0^t h_A(u)du \\ &= \exp\left(-\exp(\beta_0+\beta_1+\beta_2 x_2+\ldots\beta_p x_p)\int_0^t h_0(u)du\right) \\ &= \exp\left(-\exp(\beta_0+\beta_2 x_2+\ldots\beta_p x_p)\int_0^t h_0(u)du\right)^{\exp(\beta_1)} \\ &= S_B(t)^{\exp(\beta_1)}. \end{aligned}$$

Thus, if the hazard ratio of the treatment is 0.8, it decreases the hazard by 20% and increases survival at time t to $S_A(t) = S_B(t)^{0.8}$.

Note that by defining the survival function, we can also express the hazard function as

$$h(t) = \lim_{\delta\to 0}\frac{P(t<T<t+\delta|T>t)}{\delta} = \lim_{\delta\to 0}\frac{P(t<T<t+\delta)}{\delta S(t)}.$$

We can further re-express this term after introducing the cumulative risk function $F(t) = P(T \leq t)$, $0 < t < \infty$. Consequently, we can derive the probability density function $f(t) = \frac{d}{dt}F(t)$, which is the rate of change of the cumulative risk function. This is equivalent to the negative rate of change of the survival function, which follows from the definitions of survival function and cumulative risk function as they are complementary probabilities. Hence, the hazard function can be defined as

$$h(t) = \frac{f(t)}{S(t)},$$

which we will need further below [60].

One of the most widely used non-parametric estimators of the survival function is the product-limit or Kaplan-Meier estimator [62]. It is given by [60]:

$$\hat{S}(t) = \prod_{t_i\leq t}(1-d_i/n_i),$$

where d_i is the number of grafts, which fail at time t_i, and n_i is the number of those, which are at risk at time t_i. We can specify an alternative estimator that incorporates covariates by applying the above results. First, we define an estimate of the baseline hazard function as the ratio of deaths and the sum of estimated risk scores:

$$\hat{h}_0(t_i) = \frac{d_i}{\sum_{j\in R_i}\exp\left(x_j\hat{\beta}\right)},$$

where R_i is the set of all grafts at risk at t_i and x_j is the covariate vector of graft j. Furthermore, we set $\hat{S}_0(t) = \exp\Big(-\hat{H}_0(t)\Big)$, where $\hat{H}_0$ is the sum of the estimated baseline hazards $\hat{h}_0(t_j)$ with $t_j \leq t$. It follows that we can define the survival function estimator as

$$\hat{S}(t|x_i) = \Big(\hat{S}_0(t)\Big)^{\exp\left(x_i\hat{\beta}\right)},$$

for a covariate vector x_i of graft i.

2.1.2 Competing risks

So far, we have only considered survival for a single event. However, patients may experience different, excluding events. For example, they may die by the analysed disease, but also by an accident or any other disease. In our case, we were interested in the continued function of a transplanted kidney. However, this may be cut short due to an unrelated death of the patient. In this case, it would have been misleading to consider this as a failure of the graft, if it was still functioning at the time of death, and on the other hand, we could not expect it to continue functioning when the body had stopped working. If we were only interested in a particular outcome, in this case, the longevity of the transplant, a simple solution would have been to censor the survival time at this point. However, the fact that a patient survived with a transplanted kidney without failure until their death, is itself an important outcome we wanted to consider. In general, one of the key assumptions in censoring usually is that it is independent of the event in question, however, in a competing risk context, this assumption may often be questionable.

The following is mostly based on Moore [60]. In order to design a model incorporating competing risks, consider K distinct outcomes. Assume that every patient may experience only one outcome. We consider failure to be the ubiquitous outcome, so in our particular case, graft loss and patient death with functioning graft are the two distinct and competing causes of failure. The cumulative risk function, also called the sub-distribution function, now gives the cumulative probability of failure from cause j by time t:

$$F_j(t) = P(T \leq t, C = j) = \int_0^t h_j(u)S(u)du.$$

The cause-specific hazard h_j relates to the probability of failure at a specific time due to cause j given that a graft survived this long. Adding up all cause-specific hazards at a particular time results in the hazard function.

To model competing risks, Fine & Gray [63] introduced a sub-distribution hazard

$$\bar{h}_j(t) = \lim_{\delta \to 0} \frac{P(t < T_j < t + \delta | E)}{\delta},$$

for the event time T_j specific to cause j, given the conditional event

$$E = \{\{T_j > t\} \text{ or } \{T_i \leq t \text{ and } i \neq j\}\}.$$

Thus, the cause-specific hazard is further adapted such that it also includes grafts, which will experience another cause of death than j at a later time point than t, into the risk set. Similar to the Cox model before, Fine and Gray modelled the sub-distribution hazards for cause j as

$$\bar{h}_j(t) = \bar{h}_{0j}(t) \exp(X\beta).$$

An important measure of competing risks in given time periods is the above-mentioned cumulative incidence. In order to estimate the cumulative incidence function, that means the probability of failure due to a particular cause until some time point, let $t_1, \ldots, t_d$ be the ordered, distinct event times. In general, the hazard at time t_i can be estimated by $\hat{h}(t_i) = d_i/n_i$, where d_i is the number of events at time t_i and n_i the number of at-risk grafts at that time. For the cause-specific hazard, we restrict the former to the number of events due to cause j, d_{ij}, such that $\hat{h}_j(t_i) = d_{ij}/n_i$. Multiplying the risk of failure due to j at t_i, $\hat{h}_j(t_i)$, with the probability of being alive until just before, $\hat{S}(t_{i-1})$, results in the probability of failing due to cause j at time t_i. Finally, we get the cumulative incidence at t for cause j by adding up this probability over all prior time points:

$$\hat{F}_j(t) = \sum_{t_i \leq t} \hat{h}_j(t_i) \hat{S}(t_{i-1}),$$

where $\hat{S}(t_0)$ is typically set to 1.

Scheike & Zhang [64] introduced a generalization of the Fine and Gray model in their R package `timereg`, which we used. Consider the formulation of the cumulative incidence function

$$P_j(t|x) = 1 - \exp\left(-H_0(t) \exp\left(\beta^\intercal x\right)\right),$$

where x is some covariate vector. Scheike & Zhang [64] split up the covariate vector such that

$$\operatorname{cloglog}(1 - P_j(t|x, z)) = \gamma^\intercal x + \eta(t)^\intercal z,$$

where $\operatorname{cloglog}$ is the complementary log-log link function: $\operatorname{cloglog}(u) = \log(-\log(1-u))$,

z is a $(p+1)$-dimensional, time-varying covariate, $z = (1, z_1, \ldots, z_p)$, and x is a q-dimensional covariate.

2.1.3 Model diagnostics

In order to select the best possible models for the transplant outcome prediction tool, we used the Bayesian Information Criterion (BIC). This metric measures a model's goodness of fit and was first described in Schwarz [65]. It is defined as

$$BIC = -2\log\left(\widehat{L}\right) + p\log(n),$$

where $\hat{L}$ is the maximum value of the likelihood function of the corresponding model, p is the number of predictors, and n is the sample size. It follows that smaller BIC values indicate a better fit. For our model selection, we started with a full model, including all predictors considered relevant, and calculated the BIC. Then we fitted all models, where one of the predictors was excluded, and calculated their BIC. We moved to the model with the smallest BIC as long as it was smaller than the value of the current model. The process was repeated until there was no more improvement. After every iteration, we refitted the model with all possible data, which is relevant since some grafts had missing values for some of the predictors. Therefore, these cases could not be used for fitting models where the corresponding predictors were included. This approach is based on the algorithm described in Neumair *et al.* [66].

An alternative to excluding incomplete observations is to impute the missing values. We compared the results from the above model selection to an approach called multiple imputation by chained equations (MICE), one of the most popular imputation methods. The following description of the method is based on Azur *et al.* [67]. For MICE one assumes that missing data is missing at random, which means that the possibility that a data point is missing depends only on observed values. Because survival outcomes were observed for everybody in this study, we used MICE to impute missing covariates and not missing outcomes. MICE iteratively imputes missing covariates using the appropriate model for that covariate type using all other variables. This cycle of updating the imputed values is repeated for a set number of times to receive an imputed dataset. Finally, the whole process is repeated multiple times to generate several imputed datasets, which will later help to incorporate the uncertainty that was introduced by the imputation into the analysis results. While some publications suggest about five datasets, White *et al.* [68] advised that the number should be at least equal to the percentage of incomplete cases. A detailed introduction to the imputation and following analyses using the R

package `mice` is given by Zhang [69]. We applied our analyses to each of the datasets imputed via MICE and combined the results to compare them with the missing values approach above.

The following validation methods usually require that the modelling and validation are performed on separate datasets. After selecting suitable models in Miller *et al.* [26], we retrained the selected models on 80% of the data. We used the remaining 20% as validation data for the following metrics.

For a calibration analysis, we compared the observed risk to experience one of the three analysed outcomes, patient death, graft loss, and death with functioning graft, with the corresponding risk predicted by the model after ten years. In a well-calibrated model, these probabilities should be similar, at least for the clinically relevant ranges, that means the probabilities that occur most frequently.

A complementary method to evaluate predictive models is decision curve analysis, introduced by Vickers & Elkin [70]. The advantage of decision curve analyses is their clinical relevance in comparison to more traditional metrics evaluating calibration and discrimination, such as sensitivity, specificity, or area under the receiver operating characteristic curve (AUC), as these are limited in determining whether a model justifies its actual usability. We first define a threshold probability p_c. This probability is often determined by the physician or the graft, setting the threshold where a test is considered positive. For example, this may determine if a patient should undergo biopsy after initial cancer screening. In our case, this would translate to the assumption that the graft will experience failure up to a specific time point. We performed this analysis for all final models as well as the other outcomes, death and death with functioning graft, and set a time interval of ten years. In dependence of the threshold probability p_c, the net benefit is defined as

$$\text{Net benefit} = \frac{\text{True positives}}{n} - \frac{\text{False positives}}{n}\left(\frac{p_c}{1-p_c}\right),$$

where n is the number of grafts in the dataset. Let $\hat{p}$ denote the probability of experiencing the negative outcome, such as graft failure, based on a fitted risk model. Then the procedure for calculating the decision for further clinical action starts with selecting the threshold probability p_c. Grafts are defined positive, meaning the event has a high likelihood of occurring, if $\hat{p} \geq p_c$.

Identification of a threshold allows the calculation of the number of true and false positives, and thus the net benefit. In the kidney transplant application, true positives were defined as the number of grafts predicted to fail during the ten years past transplantation out of those who did experience mortality during this period. False positives were the

reverse, the number of grafts predicted to fail out of those who did not. The process can be repeated for a range of p_c, as well as for different models under comparison. Typically, two reference models are additionally calculated, where we assume that every prediction is either positive or negative, respectively. This translates to $\hat{p} = 1$ and $\hat{p} = 0$. In general, we should only consider clinically relevant threshold probabilities. In an optimal case, the considered model or test will have a higher net benefit than the other reference models for the relevant probabilities. For the calculation of the decision curve, we used the R package `dcurves`.

The formula above used for the calculation of the net benefit needs to be adapted if censoring or competing risks are present in the survival model. According to Vickers *et al.* [71], censoring is implemented by defining the number of true positives as $(1 - (\hat{S}(t)|\hat{p} \geq p_c)) \cdot P(\hat{p} \geq p_c) \cdot n$, where, to be clear, $\hat{S}(t)$ is the Kaplan-Meier estimator at time t, $\hat{p}$ is the predicted probability, p_c is the probability threshold, and n is the number of grafts in the dataset. Thus, instead of counting the true positives, we multiply the joint probability of a graft loss and a positive prediction for a time period of ten years with the total number of grafts. Similarly, the number of false negatives is given by $(\hat{S}(t)|\hat{p} \geq p_c) \cdot P(\hat{p} \geq p_c) \cdot n$. For competing risks, we use the cumulative incidence $\hat{F}_j(t)$ for a specific cause j instead of the Kaplan-Meier estimator, such that the true positives are set to $(\hat{F}_j(t)|\hat{p} \geq p_c) \cdot P(\hat{p} \geq p_c) \cdot n$ and the false negatives to $(1 - (\hat{F}_j(t)|\hat{p} \geq p_c)) \cdot P(\hat{p} \geq p_c) \cdot n$.

Besides calibration and decision curve analysis, we used the AUC as primary metric. We calculated the metric for the validation data using bootstrapping, meaning resampling of the data, to get confidence intervals. For five and ten years, we predicted, among other outcomes, the risk of graft loss and compared it to the actual status of the graft at the respective time. From this, we extracted the number of true positives, false negatives, false positives, and true negatives. These metrics allowed us to calculate the true positive rate, which is the ratio of true positives and the sum of true positives and false negatives. Translated to our graft survival, it is the proportion of correctly predicted losses over all actual losses. Similarly, the false positive rate is defined as the ratio of false positives and the sum of false positives and true negatives. The receiver operating characteristic curve plots the two metrics against each other, with the true positive rate as y axis and false positive rate as x axis. Thus, the AUC should be as high as possible with a maximum at 1. This is equivalent to high true positive rates, indicating that most of the graft losses were predicted, and low false positive rates, indicating that few still functioning grafts were predicted to fail. To account for censoring in the data, an inverse probability of censoring weighted estimator can be used as described by Wolbers *et al.* [72]. For these calculations, we used the R package `riskRegression` [73], which is

focused on prediction scores for survival analyses.

2.1.4 Clustering in survival analysis

In our application, kidney donors often donated two kidneys each going to a different recipient and some recipients received more than one kidney each coming from a different donor. Such clusters may indicate associations in the outcomes that cannot be identified with the available predictors. In the following, we will consider the implementation of clustering into the Cox proportional hazards model based on Moore [60].

We created clusters in Miller *et al.* [26] by considering a single recipient r and all their received grafts $g_1, \ldots, g_r$, which were part of the analysed dataset. Then we determined all donors of these grafts. If any graft of one of these donors was already included in a cluster, then we assigned $g_1, \ldots, g_r$ also to this cluster. If that was not the case, then we created a new cluster for these grafts. This was repeated for all recipients and led to a total of 18,694 clusters. The majority of clusters (n=12,665) contained two grafts, 5,408 clusters contained only one graft, and 621 clusters contained at least three grafts.

One option to consider clusters, applied to transplanted grafts in Miller *et al.* [26], is based on marginal survival models. In this approach, parameter estimates are obtained as usual from a Cox proportional hazards model under the assumption that the observations are independent. Assuming that the model is correctly specified, Lee *et al.* [74] showed that this estimator is consistent and asymptotically normal, but its covariance matrix estimate has to be corrected. Let $\hat{V}$ be the usual covariance estimate under the independence assumption and G be the number of clusters. We define for graft i in cluster k, $i = 1, \ldots, n_k$, $k = 1, \ldots, G$, a score residual:

$$s_{ij} = \delta_{ki}(x_{ki} - \bar{x}(t_{ki})) - \sum_{t_u \leq t_{ki}} (x_{ki} - \bar{x}(t_{ki})) \exp(x_{ki}\beta)(\hat{H}_0(t_u) - \hat{H}_0(t_{u-1})),$$

where δ_{ki} is the event indicator, which means $\delta_{ki} = 0$ for censoring and $\delta_{ki} = 1$ for loss of graft i, t_{ki} is the event time, x_{ki} are the covariates, and $\bar{x}(t_{ki})$ are the expected values of the covariates at this time, a weighted sum of the covariate values for grafts at risk at t_{ki}. $\hat{H}_0$ is the estimated baseline cumulative hazard. Assuming a model with p covariates, then s_{ki} is a $1 \times p$ matrix. A cluster-adjusted standard error for the coefficient estimates is then defined by $diag(V^*)^{1/2}$, where the adjusted covariance matrix $V^* = \hat{V}C\hat{V}$, with $C = \sum_{k=1}^{G} \sum_{i=1}^{n_k} \sum_{l=1}^{n_k} s_{kj}^{\intercal} s_{kl}$ and G as number of clusters, is used instead of the usual covariance estimate. For a detailed derivation see the initial introduction of this approach in Lee *et al.* [74] or Klein & Moeschberger [75]. In R, this approach is implemented by expanding the formula of the function `coxph` with a `cluster()` term.

Another approach to account for clusters adapts the hazard function with a common factor for each graft called frailty. It can be considered a random effect. For this method, let ω_i be the frailty for all grafts in the ith cluster. Let (t_i, δ_i, x_i) be the ith observation. Consider the likelihood function:

$$\begin{aligned} L(\beta|x_i) &= \prod_{i=1}^{n} f(t_i,\beta)^{\delta_i} S(t_k,\beta)^{1-\delta_i} \\ &= \prod_{i=1}^{n} h(t_i,\beta)^{\delta_i} S(t_k,\beta) \\ &= \prod_{i=1}^{n} \left(h_0(t_i)\exp(x_i\beta)\right)^{\delta_i} \exp(-H_0(t_i)\exp(x_i\beta)) \end{aligned}$$

using $h(t) = f(t)/S(t)$, $h(t,\beta) = h_0(t_i)\exp(x_i\beta)$, and $H_0(t_i) = -\int_0^{t_i} h_0(v)dv$. Then incorporate the frailty term into the hazard function for the ith graft in the kth cluster by defining

$$h(t_{ki}|\omega_i, x_{ki}, \beta) = h_0(t_{ki})\omega_i \exp(x_{ki}\beta).$$

The frailty depends on the G clusters and can be modelled via some distribution like the gamma distribution: Often, it is assumed that $\omega \sim g(\omega|\theta)$ follows a gamma distribution, such that the joint likelihood for the ith graft in cluster k can be written as

$$L(\beta,\theta|\omega_k, t_{ki}, \delta_{ki}, x_{ki}) = g(\omega_k|\theta)(h_0(t_{ki})\omega_i \exp(x_{ki}\beta))^{\delta_{ki}} \exp(-H_0(t_{ki})\omega_k \exp(x_{ki}\beta))).$$

As an alternative to using the gamma distribution, we can also write $\omega_k = \exp(\sigma u_k)$, where u_k has a standard normal distribution, such that random and fixed effects are on the same level:

$$h(t_{ki}|\omega_i, x_{ki}, \beta) = h_0(t_{ki})\exp(x_{ki}\beta + u_k\sigma).$$

Finally, the expectation-maximization algorithm, a well-known iterative process to find maximum likelihood estimates described for example in Dellaert [76], can be used to estimate the parameters. In R, the frailty model can be implemented by expanding the formula of the function `coxph` with a `frailty()` term, which uses by default the gamma distribution.

There are some reviews on the decision to include cluster effects in the model, however, previous publications often considered multi-centre clinical trials in which a clinic represented a cluster, which is quite different from the clusters we considered. Nonetheless, Glidden & Vittinghoff [77] summarize several reviews by deriving that fixed effects should be chosen if inference is only intended to apply for the selected centres and for more general inferences, random effects are preferred. Yet, Snavely [78] argues

that frailty models should be used for centre-specific interpretations, while the marginal model should be used if population-averaged interpretation is of interest. In the first study, cluster sizes between 2 and 20 and a total sample size of 100 and 400, in the latter, a cluster size of 10-50 and a sample size of 500 were simulated. Therefore, these results have very limited applicability to our situation with very small clusters and much larger sample sizes. Further simulation studies are urgently needed to determine the relevance of cluster adjustments and which method is most appropriate in comparable survival analyses.

2.2 Published COVID-19 time series models

The COVID-19 pandemic catalysed statistical methodological research to reproduce and simulate the pandemic's development, with a primary focus on the modelling and prediction of infection or death. Some of the more highly publicized models are listed in Table 2.1. In the following, we present a diverse selection of early publicized COVID-19 models to understand core ideas that dictate state-of-the-art model development for infectious diseases.

2.2.1 Models of early developments in China

Wu *et al.* [86] were one of the first to forecast the spread of COVID-19 in China and internationally. In order to simulate the Wuhan epidemic, the authors employed a susceptible-exposed-infectious-recovered (SEIR) model, where $S(t)$, $E(t)$, $I(t)$, and $R(t)$ are the numbers of susceptible, exposed, infectious, and recovered individuals at time t. SEIR models are one of the major schools of methods modelling infectious diseases. Usually based on ordinary differential equations, they are deterministic and derive from the more basic susceptible-infectious and susceptible-infectious-recovered models [94]. In the susceptible-infectious model, the population is partitioned into two distinct groups. The union of these disjoint groups covers the whole population. One group, the infectious, is infected with respect to a specific disease and is capable of infecting individuals of the other group, the susceptibles. The counts of these groups are variables and the flow between them is typically modelled by differential equations. For the SEIR approach, the partition of the population is extended from two to four compartments. Next to the susceptible and the infectious group, there are now also the exposed, who are infected but not yet infectious, and the removed or recovered individuals. A more detailed introduction to the concept can be found in Murray [94].

Publication	Disease	Main statistic	Model framework	Model details	Location	Time range
Schneble *et al.* [79]	COVID-19	Daily deaths	Generalized additive mixed model	Flexible statistical models with smooth spatial components and a district-specific Poisson process	German districts	Mar 26 - May 13, 2020
Wang *et al.* [80]	COVID-19	Daily infections	Statistical survival-convolutional model	Transmission rate modelled as non-negative piecewise linear function with knots placed at meaningful event times	China / South Korea / Italy / USA	Jan 20 - Feb 4 / Feb 15 - Mar 4 / Feb 20 - Apr 29 / Feb 21 - May 1, 2020
IHME [81]	COVID-19	Cumulative death rate	Mixed effect non-linear regression	Cumulative death rate as parametrised Gaussian error function; state-specific social distancing covariates	13 regions with peak deaths in China, Italy, Spain, USA	Until Apr 14, 2020
Mercker *et al.* [82]	COVID-19	Infection rate	Generalized additive mixed model (GAMM)	GAMM of time series of districts with federal state as random intercept, autoregressive structure with order 1; NegBin distribution	Germany	Until Apr 13, 2020
Guenther *et al.* [83]	COVID-19	Daily infections	Hierarchical Bayesian	Poisson and Binomial distribution for cases and their delays	Bavaria (Germany)	Mar 1 - Apr 9, 2020
Woody *et al.* [84]	COVID-19	Daily death rate	Generalized linear model	Similar to IHME model but daily death rates as mixed effect NegBin generalized linear model	USA	Until Apr 3, 2020
Flaxman *et al.* [85]	COVID-19	Daily deaths	Semi-mechanistic Bayesian hierarchical model	Deaths as sum of past infections weighted by probability of death; expected number of deaths expected to follow NegBin; infections as sum of infections in previous days, weighted by discretized generation distribution	11 European countries	Until Mar 28, 2020
Wu *et al.* [86]	COVID-19	Cumulative infections	SEIR	Modelling change in compartments based on prior numbers and disease periods	mainly China	Dec 31 - Jan 28, 2020
Charu *et al.* [87]	Influenza	Weekly infections	Partial likelihood of spatial locations	Semi-parametric model for the conditional intensity giving the hazard of infection in a specific location based on proximity to other locations	USA	2002 - 2010
Mahsin *et al.* [88]	Influenza	Daily infections	Geographically dependent discrete-time individual-level model	Conditional autoregressive to capture spatially structured latent covariates	Calgary (Canada)	Oct 25 - Nov 14, 2009
Lowe *et al.* [89]	Dengue fever	Monthly infections	Generalised linear mixed model	Count of cases as NegBin	Brazil	2001 - 2009
Meyer *et al.* [90]	Meningococcal disease	Weekly infections	Additive model of epidemic and endemic parts	Endemic: piecewise constant function on spatiotemporal grid; epidemic: the sum of factorized effects of marks, elapsed time, and relative location	German districts	2002 - 2008
Paul *et al.* [91]	Influenza with meningococcal disease / Influenza	Weekly infections	Additive model of epidemic and endemic parts	Epidemic: autoregression on previous cases with NegBin; endemic: harmonics with incidence levels dependent on subgroups	Germany / USA	2001 - 2006 / 1996 - 2006
Hefley *et al.* [92]	Chronic waisting in white-tailed deer	Overall infections	Generalized linear mixed model for spatial processes	Spatial process in a hierarchical model	Wisconsin (USA)	2002
Malesios *et al.* [93]	Sheep pox	Weekly farms with infections	Spatio-temporal Bayesian model	Extend spatio-temporal model with Ornstein-Uhlenbeck process to incorporate features of epidemic data	Greece	1994 - 1998

Table 2.1: Selection of infectious disease models. NegBin corresponds to the negative binomial distribution, GAMM to generalized additive mixed model, SEIR to susceptible-exposed-infectious-recovered model.

Wu *et al.* [86] modelled the spread of the virus as a function of the changes in S, E, and I dependent on day t via derivatives as follows:

$$\frac{\mathrm{d}S(t)}{\mathrm{d}t} = L_{I,W} + L_{C,W}(t) - \frac{S(t)}{N}\left(\frac{R_0}{D_I}I(t) + z(t)\right) - \left(\frac{L_{W,I}}{N} + \frac{L_{W,C}(t)}{N}\right)S(t),$$
$$\frac{\mathrm{d}E(t)}{\mathrm{d}t} = \frac{S(t)}{N}\left(\frac{R_0}{D_I}I(t) + z(t)\right) - \frac{E(t)}{D_E} - \left(\frac{L_{W,I}}{N} + \frac{L_{W,C}(t)}{N}\right)E(t),$$
$$\frac{\mathrm{d}I(t)}{\mathrm{d}t} = \frac{E(t)}{D_E} - \frac{I(t)}{D_I} - \left(\frac{L_{W,I}}{N} + \frac{L_{W,C}(t)}{N}\right)I(t).$$

R is not explicitly defined in the publication, but it would comprise at least the recovered individuals leaving the group of infected, expressed in the third equation as the factor $\frac{I(t)}{D_I}$. Above equations require specification of nine parameters, D_E, D_I, R_0, N, $z(t)$, $L_{W,I}$, $L_{I,W}$, $L_{W,C}(t)$, and $L_{C,W}(t)$, which we now define in turn.

N is the catchment population size of the Wuhan Tianhe International Airport at Wuhan. The authors set the passenger numbers to the numbers of previous years, where $L_{W,I}$ was the daily average of international outbound air passengers from mainland China, $L_{I,W}$ the international inbound air passengers, $L_{W,C}(t)$ the daily number of all domestic outbound travellers, and $L_{C,W}(t)$ the domestic inbound travellers. Figure 2.2 illustrates the intensity of outbound passenger numbers and shows that regional travel inside the country is very heterogeneous and highlights the importance in comparison to international travel.

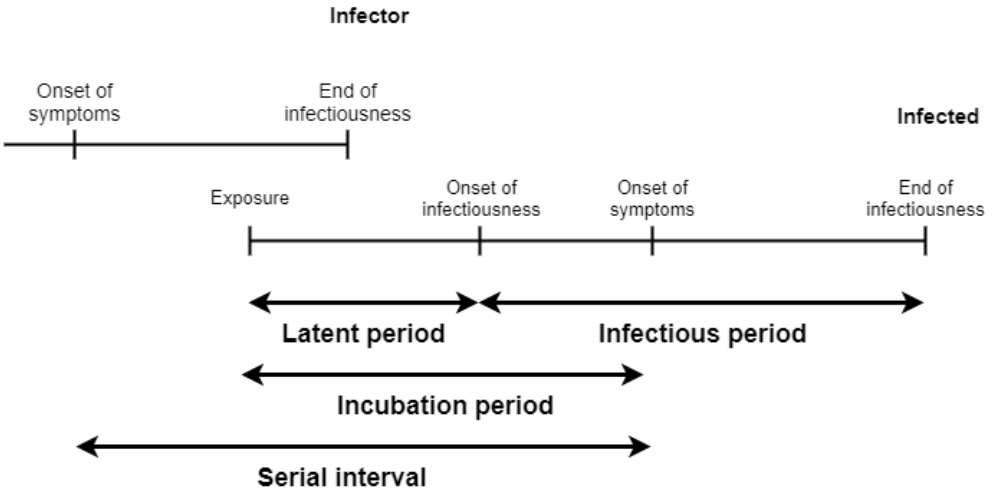

Figure 2.1: Diagram of time intervals in infectious diseases (based on Lipsitch *et al.* [95] and Porta [96]).

The basic reproductive number, R_0, is the average number of infected people by a

single infectious person, if the whole population is assumed to be still susceptible to the virus. In contrast to the other parameters, R_0 is not taken from literature but modelled as discussed further down.

The components D_E and D_I are the mean latent and infectious periods, defined as the time range between exposure and beginning of infectiousness and between onset and end of infectiousness, respectively (see Figure 2.1). They assumed that the latent period equals the incubation period and the infectious period the serial minus the latent period, and used time ranges from the literature from the previous two major coronaviruses SARS-CoV and MERS-CoV.

The zoonotic force of infection, $z(t)$, was set in the baseline scenario to 86 cases per day until January 1, 2020, and then 0 thereafter, as January 1 was the day the Huanan seafood wholesale market, the most probable index source, was closed.

The derivatives represent the change in each of the compartments, S, E, and I, over time. Each compartment has incoming and outgoing flows. All three compartments have one outgoing number which is the count of departing passengers, $L_{W,I}$ and $L_{W,C}(t)$, multiplied by the proportion of the respective compartment in the local population. We can see in the first equation that the number of susceptible individuals increases by the incoming passengers. The number of new infections, either by zoonotic force or by contact with other infected decreases the number of susceptible. The zoonotic force is fixed as described above, while the contact infections are determined by the number of currently infectious individuals factored by the strength, R_0, over the duration, D_I, of the infectiousness.

These newly infected are then counted towards the exposed group, visible in the second equation. People in the exposed group either travel outward or stay in the area and become infectious after the latent period, thereby counting towards the infectious group then.

Decreasing summands of the infectious group are given by the outflow of travellers and the people who are not infectious anymore. For this, the infectious cases are divided by the infectious period.

Wu *et al.* [86] then modelled the international case exportation by a non-homogeneous process with rate

$$\lambda(t) = \frac{L_{W,I}}{N}(E(t) + I(t)),$$

assuming that travel behaviour was not affected by the disease. Thus, the case exportations are assumed to be a Poisson process with a mean dependent on time and defined as the number of outbound travellers multiplied by the proportion of exposed and infectious individuals in the local population, thereby assuming that the proportion

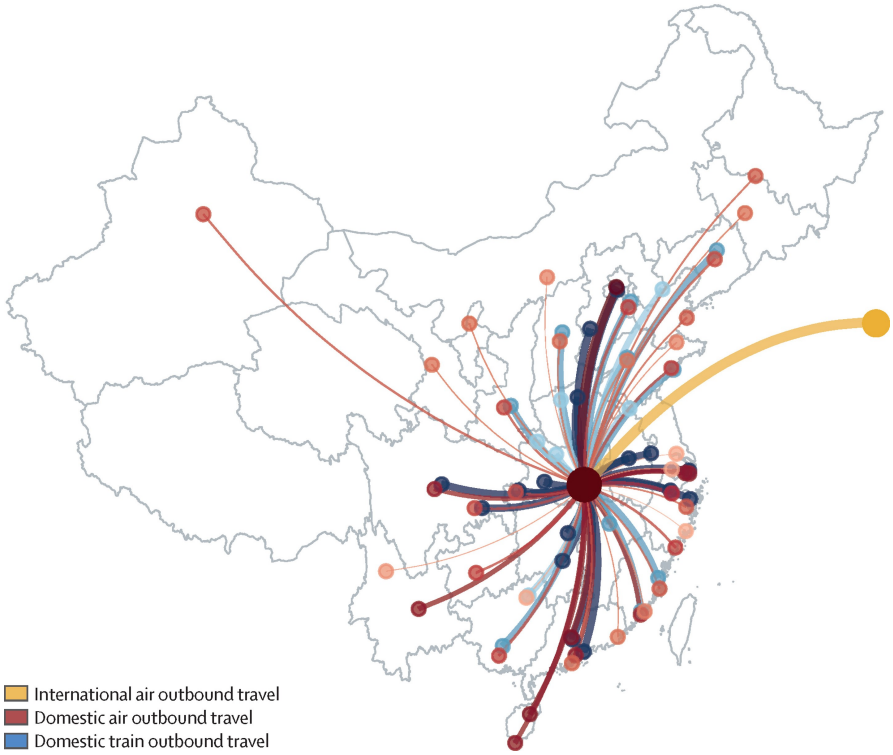

Figure 2.2: Major routes of outbound air and train travel originating from Wuhan during the Spring Festival travel season, 2019. Darker and thicker edges represent greater numbers of passengers (from Wu *et al.* [86]).

of cases in the local population and for the travellers is equal. As seen in the equations above, these two groups have the parameter basic reproductive number R_0. They applied Markov Chain Monte Carlo methods with Gibbs sampling and non-informative flat prior distributions to estimate R_0 assuming the model and mobility and flight data (Figure 2.2).

Shortcomings of the publication included that many assumptions were required due to the sparse data. For example, the mean serial interval of COVID-19 was calculated to be 3.96 days based on 468 COVID-19 transmissions between January 21 and February 8, 2020 [97], which was lower than the respective period of SARS-CoV. Furthermore, some assumptions could not be projected into the future such as the unaffected travel behaviour. Similarly, the model assumed that all cases showed symptoms at some point, but the literature reported from 4 to 42 percent asymptotic cases among positive tested individuals [98, 99]. Other assumptions could also be unreliable, such as the zoonotic mechanism.

There were a number of other COVID-19 publications utilizing the SEIR modelling

framework [100–105]. However, SEIR models are not tailored for the implementation of dynamic changes, such as the introduction of non-pharmaceutical interventions (NPI), which can have a big impact on the reproductive rate. Therefore, many SEIR models have failed to forecast downward trends in the number of deaths or infected [81]. Another limitation is their reliance on the number of reported infections, which is in many cases not comparable due to differences in testing systems between regions as well as changes over time.

2.2.2 The IHME model

The Institute for Health Metrics and Evaluation (IHME) 2020 group also took a deterministic approach but instead of using the SEIR model, modelled the cumulative death rate as a function of time on a regional level. They modelled deaths rather than infections based on their higher accuracy, and only considered locations with a death rate greater than $\exp(-15) \approx 0.31$ per million.

They assumed that cumulative death rates follow a Gaussian error function:

$$D(t;\alpha,\beta,p) = \frac{p}{2}\left(1 + \frac{2}{\sqrt{\pi}}\int_0^{\alpha(t-\beta)} \exp\left(-\tau^2\right) d\tau\right), \tag{2.1}$$

where p denotes the location-specific maximum death rate, t the time since the death rate exceeded $\exp(-15)$, β the location-specific time axis shift parameter that indicated the time at which the rate of change of the cumulative death rate D is maximal and α is a location-specific growth parameter, which can be interpreted as the speed of the infection.

An ensemble of three mixed-effects models was applied to the log of the death rate that differed by the definition of social distancing applied to regions from China, Italy, Spain, and USA, which were assumed to have reached peak deaths by April 14, 2020. For each model, there were two variants, a short-term to explain existing data and a long-term for forecasting. A predictive validity framework that analysed errors in predicting out-of-sample observations was used to calculate uncertainties of models.

Jewell *et al.* [106] noted that a disadvantage of the model was the strong assumption that epidemic curves follow a normal distribution. Another assumption of the model was that mortality rates were influenced by introduction dates of social distancing measures in a similar manner across different countries with differences modelled by random effects. Although a uniform response may apply for early epidemic curves, it can be expected to deviate with time and by country. Analyses of mortality in US cities during the 1918 influenza pandemic showed that not only the timing of implementation of

non-pharmaceutical interventions influenced the death rate, but also the timing of their relaxation [107, 108]. Figure 2.3 from Markel *et al.* [107] shows peaks in mortality rate following the end of bans and shows different distributions in accordance with different interventions. Another issue was that early on, when only few complete mortality curves

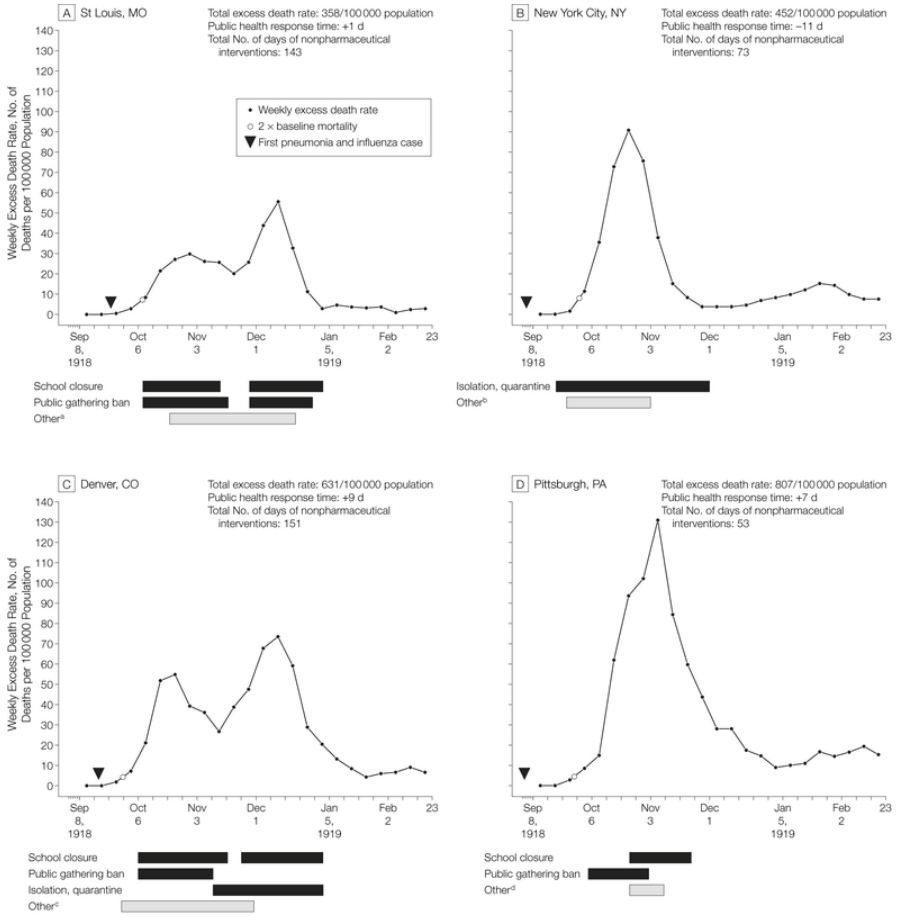

Figure 2.3: Weekly excess death rates during the 1918 influenza pandemic from September 8, 1918, through February 22, 1919 (from Markel *et al.* [107]).

were observed, estimated parameters, such as curve inflection times, may have been unduly influenced by unrepresentative regions and policies supplying the data.

The model offered a broad range of uncertainty, especially considering that some sources, such as data and model structure, were not incorporated. For example, projected cumulative US deaths based on data from mid-April plateaued after about three weeks at 60,000 counts with a 95% confidence interval ranging from 34,000 to 140,000 until the end of the prediction period around the end of June. Furthermore, predictions sometimes changed drastically when new input data was provided, without

any apparent reason. Marchant *et al.* [109] also found that 49-73% of observed death counts fell outside a 95% prediction interval based on an early version of the model in contrast to the expected 5%.

Other approaches have also modelled death counts. Atkeson *et al.* [110] compared a stylized version of the IHME model to a SIR model. Schumacher *et al.* [111] utilized a nonlinear mixed-effects model for the death numbers in Latin American countries, and Schüttler *et al.* [112] employed a Gaussian function of time for the logarithm of daily death rates in 25 countries.

2.2.3 The University of Texas model

Woody *et al.* [84] shifted from the cumulative to the daily death rate, thus considering the derivative of the key equation from the IHME model (2.1):

$$D'(t) = \frac{p\alpha}{\sqrt{\pi}} \exp\left(-\alpha^2(t-\beta)^2\right).$$

with a different parametrization. The daily death rate λ_{it} was modelled as

$$\mathrm{E}(y_{it}) = \lambda_{it} = \kappa_{it} \exp\left(-\frac{\omega_{it}}{2}(t-\mu_{it})^2\right),$$

where t was the number of days since deaths crossed the threshold of 3 per 10 million, y_{it} is the observed number of deaths in an area i at time t, $\kappa = \frac{p\alpha}{\sqrt{\pi}}$ the maximum daily expected death rate, $\mu = \beta$ the day on which the expected death rate achieved its maximum, and $\omega = 2\alpha^2$ a steepness parameter for the rise or fall of the death rate. The authors assumed the log death rate varied quadratically with time t:

$$\log \lambda_{it} = \log \kappa_{it} - 0.5\omega_{it}(t-\mu_i)^2 \equiv \beta_{i0} + \beta_{i1}t + \beta_{i2}t^2,$$

with $\omega = -2\beta_2$, $\mu = \frac{\beta_1}{2\beta_2}$, and $\kappa = \exp\left(\beta_0 - \frac{\beta_1^2}{2\beta_2}\right)$.

At this point, the authors moved away from an assumed Normal distribution for the number of deaths y_{it} and assumed instead the more appropriate negative binomial distribution for counts. One strong constraint of the Poisson distribution, a typical count model distribution, is that expected value and variance are equal. If we consider a Poisson model for count data $y \sim \mathrm{Poisson}(\phi), \phi > 0$, and assume that the mean ϕ is a Gamma distributed random variable, $\phi \sim \mathrm{Gamma}(r, \frac{r}{\lambda})$, we obtain the negative binomial distribution, $\mathrm{NegBin}(\lambda, r)$. Therefore, the negative binomial has the major advantage of higher flexibility over the Poisson distribution. The $\mathrm{NegBin}(\lambda, r)$ distribution is determined by two parameters, the mean λ and the overdispersion parameter r. The

implied variance is $\lambda(1 + \frac{1}{r}\lambda)$. For $r \to \infty$, variance approaches the mean and the distribution approaches again a Poisson distribution.

Normally distributed random effects were incorporated to yield the full hierarchical specification of the University of Texas (UT) model:

$$\begin{aligned} y_{it} &\sim \mathrm{NegBin}(\lambda_{it}, r), \\ \log \lambda_{it} &= \log N_i + \beta_{it,0} + \beta_{it,1} t + \beta_{it,2} t^2, \\ (\beta_{it,0}, \beta_{it,1}, \beta_{it,2})^T &= \mu + \Gamma x_{it} + \nu_i, \\ \nu_i &\sim \mathcal{N}(0, \Sigma), \end{aligned}$$

where x_{it} was a vector of social distancing predictors and N_i the population of area i as offset to obtain the rate of deaths in the respective area. The authors included extensions to the model, including a fixed effect for weekend days in the linear expression for the log of the death rate λ to incorporate under-reporting of numbers on weekends. The authors fit the model using Markov Chain Monte Carlo in a Bayesian approach with prior distributions β via the `R` package `rstanarm` [113].

The expectation of the death rate was quadratic similar to the Gaussian kernel of the IHME model. The UT team implemented additional mobility data to quantify movement and social distancing, similar to the later IHME models. GPS traces of mobile phones provided surrogates of visitor density in public and working spaces. In detail, mobile phone data described the median time people spent at home, the number of people spending full time at their usual place of work, and the total per capita visitation counts of specific types of public places. This is visualized in Figure 2.4, which shows that the mobility indicators started dropping off at the beginning of March with colleges and schools showing the earliest decline and grocery stores first increasing shortly and then also declining at the end of March.

The original IHME model had a statistical flaw in the assumptions that exaggerated certainty in forecast results. Namely, in modelling cumulative death rates their procedure erroneously assumed successive model errors were independent. But as the estimation goal was cumulative, values on one day depended on values of previous days. In contrast, the UT group modelled daily death rates, however, they did not apply an autoregressive error structure to account for the likely dependence between days.

Other publications employed comparable techniques in the modelling of COVID-19. Millett *et al.* [114] built a zero-inflated negative binomial regression model to determine racial disparities in US counties regarding cases and deaths. Guenther *et al.* [83] used a Bayesian hierarchical model for fitting the number of cases with a negative binomial

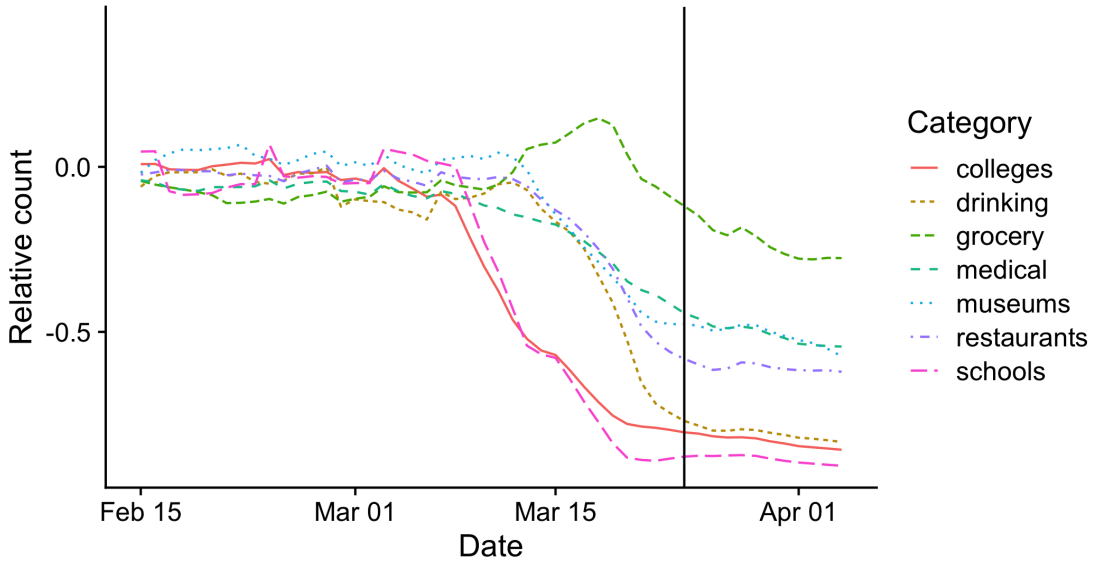

Figure 2.4: Visitation patterns in Texas visualized as relative changes to pre-pandemic baselines for various points of interest (from Woody *et al.* [84]). The vertical line showing when the state death rate reached 3 per 10 million residents.

distribution in order to nowcast, i.e. correct daily case counts for occurred-but-not-yet-reported events, in districts in Bavaria, Germany. Applying the same framework, Neelon *et al.* [115] compared the temporal effect on infection and death rates between US counties with respect to their social vulnerability. Coker *et al.* [116] used a negative binomial model to suggest a positive association between the overall excess mortality due to COVID-19 and air pollution in municipalities in Northern Italy.

2.2.4 The Imperial College model

The Imperial College COVID-19 Response Team model inferred bounds on infection numbers, case detection probabilities, and reproduction numbers R_t over time for individual countries [117]. In order to achieve this, they fit a Bayesian model of the mechanics of the infection cycle to the observed deaths for eleven European countries based on data up to March 28, 2020. The death count was taken as a more reliable dependent variable compared to the reported number of infected. An update to the first publication expanded the original model to include population saturation effects, prior uncertainty on the infection fatality ratio, better balancing on intervention effects, and partial pooling of the lockdown intervention effect [85]. However, one key property of the model remained, namely, that the authors did not derive a number of parameters empirically but rather took estimates from the literature.

They assumed

$$D_{t,m} \sim \text{NegBin}\left(d_{t,m}, \Phi\right),$$
$$\Phi \sim \mathcal{N}^{+}(0, 5),$$

where $D_{t,m}$ are the daily deaths attributed to COVID-19 for the days $t \in 1, \ldots, n$ and countries $m \in 1, \ldots, p$. Parameter $d_{t,m}$ is the mean and Φ the overdispersion parameter assumed to follow a half normal distribution. To avoid deaths resulting from non-locally acquired infections, the authors only included deaths from days after regions had cumulatively observed at least ten deaths.

In order to link the number of deaths with the number of infected, the authors utilized two quantities. One was the infection-fatality-ratio, ifr^{*}, and the other, was the time from infection to death π. Then, country-specific expected deaths for specific days and age groups were modelled as

$$\begin{aligned}
d_{t,m,a} &= ifr^{*}_{m,a} \sum_{\tau=0}^{t-1} \pi_{t-\tau} c_{\tau,m}, \\
ifr^{*}_{m} &\sim ifr_{a} \cdot \mathcal{N}(1, 0.1), \\
ifr_{a} &= \frac{AR_{50-59}}{AR_{a}} ifr'_{a}, \\
AR_{a} &= c_{a}/N_{a} \\
\pi_{s} &= \begin{cases} \int_{0}^{1.5} \pi(\tau)\, \mathrm{d}\tau & \text{for } s = 1 \\ \int_{s-0.5}^{s+0.5} \pi(\tau)\, \mathrm{d}\tau & \text{for } s = 2, 3, \ldots \end{cases} \\
\pi &\sim \text{Gamma}(5.1, 0.86) + \text{Gamma}(18.8, 0.45).
\end{aligned}$$

The expected number of deaths for country m on day t was modelled as the sum of past infections weighted by their probabilities of death, which depended on the number of days since infection. The parameter $c_{\tau,m}$ was the number of new infections on day τ and location m to be modelled as shown later. In order to incorporate uncertainty in the estimate of the country-specific mean infection-fatality ratio ifr_{m}, the authors incorporated additional noise around the mean and created ifr^{*}_{m}. Parameter ifr', which assumed homogeneous attack rates across age groups, was identified from the literature. The attack rate is a synonym here for the infected proportion of the population. ifr' was adapted by age-group-specific attack rates to obtain ifr_{a}. The term c_{a} was assumed to be the number of infections, N_{a} the underlying size of the population, and AR_{a} the attack rate, all three specific to the age group a. The time from infection to death π

was defined as the sum of the incubation period and the time between the onset of symptoms and death, assumed Gamma distributed with parameters taken from the literature. Its density $\pi(\tau)$ was used to discretize it to π_s.

A discrete renewal process was used to model the true number of infected c, thus the number of infections on day t in country m was given by

$$\begin{aligned} c_{t,m} &= S_{t,m} R_{t,m} \sum_{\tau=0}^{t-1} c_{\tau,m} g_{t-\tau}, \\ S_{t,m} &= 1 - \frac{\sum_{i=1}^{t-1} c_{i,m}}{N_m}, \\ g &\sim \text{Gamma}(6.5, 0.62). \end{aligned}$$

The time between when a person gets infected and when they subsequently infect other people was given by the density $g(\tau)$ with the generation distribution g, assumed to be equal to the serial interval distribution and Gamma distributed. Furthermore, N_m was the population of country m, and $S_{t,m}$ an adjustment factor accounting for herd immunity reducing the number of newly infected. The model assumed that reinfection was not possible over the considered time frame and did not account for infected who died before infecting others. The parameter $R_{t,m}$ was the country-specific time-varying reproduction number, scaled from a baseline prior $R_{0,m}$ by piecewise constant functions driven by NPI.

Country- and time-specific reproductive numbers were modelled with multiplicative effects of interventions:

$$\begin{aligned} R_{t,m} &= R_{0,m} e^{-\sum_{k=1}^{6} \alpha_k I_{k,t,m} - \beta_m I_{5,t,m}}, \\ \beta_1, \ldots, \beta_M &\sim \mathcal{N}(0, \gamma), \\ \gamma &\sim \mathcal{N}(0, 0.2), \\ \alpha_k &\sim \text{Gamma}(1/6, 1) - \frac{\log(1.05)}{6}. \end{aligned}$$

The indicator functions $I_{k,t,m}$, $k \in \{1, \ldots, 6\}$ equalled one if intervention k was in place in country m at time t, otherwise zero. The six interventions considered were the closure of schools and universities ($k = 1$), self-isolation if ill ($k = 2$), banning of public events ($k = 3$), any intervention in place ($k = 4$), partial or complete lockdown ($k = 5$), and encouragement of social distancing ($k = 6$). The fourth covariate ($k = 4$) indicated only whether any of the five interventions were in effect. Figure 2.5 shows that the introduction dates of interventions were in some cases very different and in other cases

very similar between the countries. The effects of interventions were assumed to be multiplicative. The parameter β_m was the country-specific random effect of a lockdown, while all impacts α_k were shared between the countries. Individual parameters α_k were chosen to have a probability that individual interventions do not reduce $R_{t,m}$, $P(\alpha_k < 0)$, near 48% and that the joint effect of all α_k on the probability be uniformly distributed on $[0, 1.05]$ if all interventions are in place.

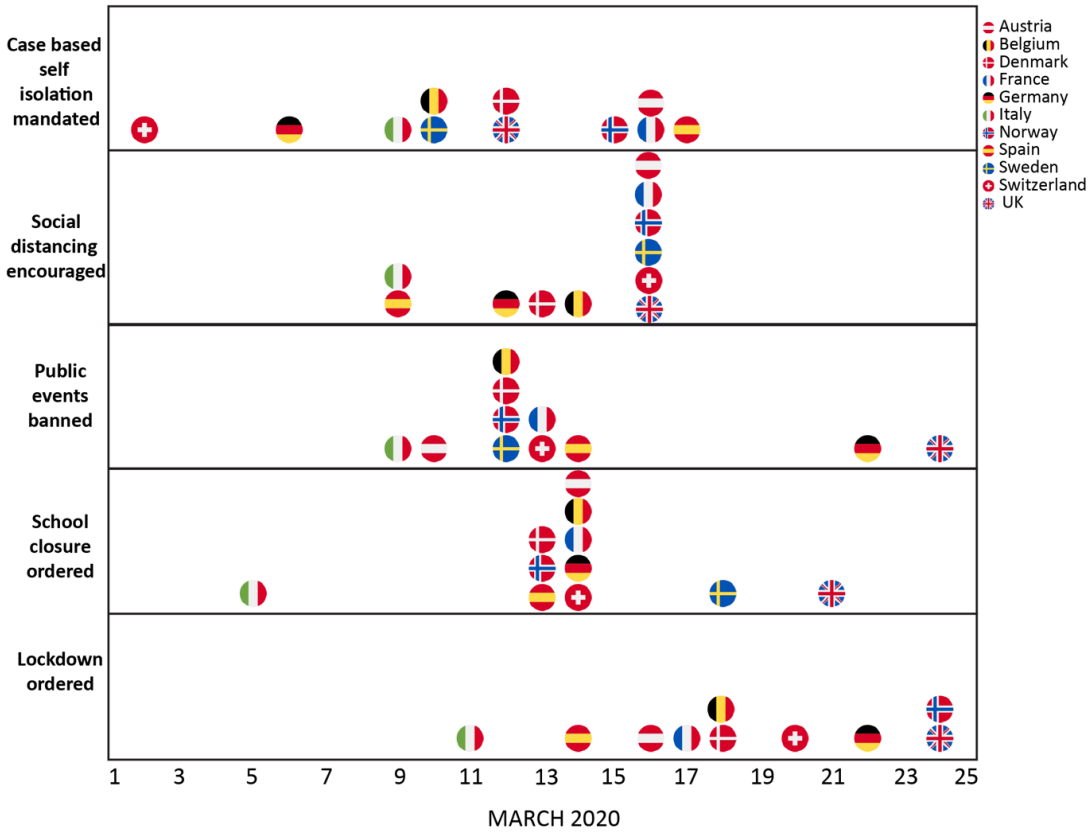

Figure 2.5: Introduction dates of interventions (from Flaxman *et al.* [117]).

A focus of the Imperial College model was the reproduction number R, which is the average number of infections generated directly by one infected case. The effective reproduction number at time t, R_t, generalizes the basic reproduction number, R_0, defined as the expected number of infections generated by one case, in the sense that R_0 assumes that all of the population is susceptible whereas R_t does not. A value of $R_t < 1$ indicates that new infections are decreasing, implying control of the epidemic, and $R_t > 1$ that they are increasing until herd immunity is achieved and the epidemic

ultimately declines. Lipsitch *et al.* [95] formulated the basic reproductive number as $R_0 = kbD$, where k is the number of contacts an infectious person has per unit of time, b is the probability of infection per contact, and D is the average infectiousness duration. The effective reproductive number was then defined as $R_t = R_0 x$, where x is the susceptible proportion of the population.

The overall approach of reliance on key values from external studies allowed to make the individual parameters more robust, yet, it reduced the model's flexibility and made it more reliant on the infection process. Furthermore, due to the different stages of the epidemic curves in the incorporated countries, higher counts of death and earlier interventions had an increased impact on the model, which may have introduced bias. However, this disadvantage applied to all models integrating different administrative regions.

Other publications also used the presented method or followed similar approaches. Delius *et al.* [118] analysed the infection-fatality-ratio in European data by adapting the Imperial College model. Olney *et al.* [119] used it to measure the effect of social distancing interventions based on data from the US. Brauner *et al.* [120] went even further and applied a similar framework to analyse interventions for 41 countries. Saez *et al.* [121] used generalized linear mixed models to monitor the development and effect of interventions on COVID-19 infections and deaths in the context of socio-economic and demographic factors.

2.3 Modelling options for longitudinal count data with autoregression

In the following, we discuss modelling options for count data. We compare Bayesian approaches with frequentist methods as a basis for the modelling decisions made in later sections.

2.3.1 Simple Poisson Model

We first considered a simple autoregressive Poisson distribution as base model to infer infection numbers on one day given the count of the previous day:

$$\begin{aligned} I_t|\lambda_t &\sim Poisson(\lambda_t) \quad \forall t \in \{1, \dots, n\} \\ \log(\lambda_t) &= \beta_0 + \alpha \log(I_{t-1} + 1), \end{aligned} \tag{2.2}$$

where I_t are the infections on day t and n is the number of observed days. This model is similar to integer-valued generalized autoregressive conditional heteroscedasticity (INGARCH) models often used in finance mathematics [122]. However, we used a log link function extension, which considers the exponential growth rates and enables a more intuitive generalization with external factors. The model will later be extended as this simple specification is not yet able to simulate more than a stable situation with constant dynamics as demonstrated in Figure 2.6. Nonetheless, we will discuss this basis in detail. The parameter combinations in the figure will later be considered for simulation studies.

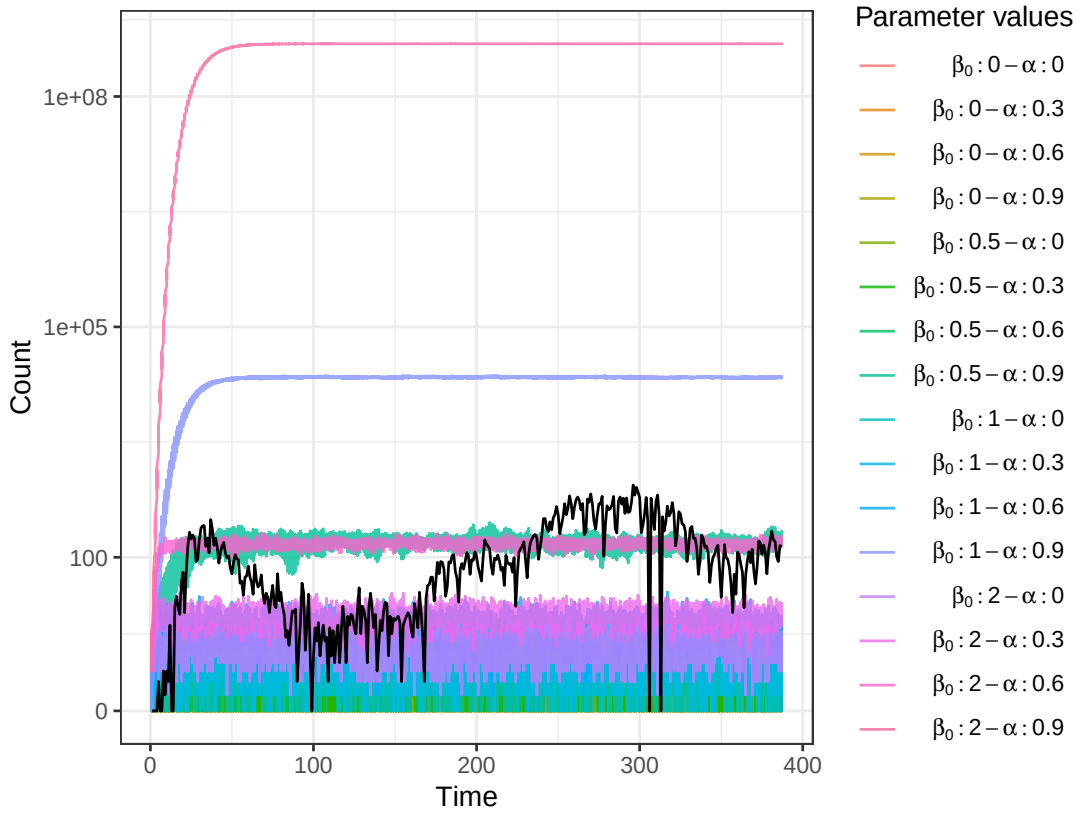

Figure 2.6: Time series of daily COVID-19 infections in Munich from 2020-02-24 to 2021-03-16 (black) in comparison to Poisson models (coloured) with different parameter combinations each with 10 simulations.

We utilized Markov Chain Monte Carlo (MCMC) with the Metropolis algorithm to sample from the posterior $f(\theta|I)$, where $I = (I_0, \ldots, I_n)$ and $\theta = (\alpha, \beta_0)$, to fit the hierarchical model. The procedure is briefly illustrated in Algorithm 1. The individual steps, particularly the key calculation of the acceptance probability and its log transformation, are discussed below.

Algorithm 1 Metropolis sampler

procedure METROPOLISSAMPLER($iterations$, $startvalues$, Σ, $data$)
  $\theta_{\text{old}} \leftarrow \theta^{(1)} \leftarrow startvalues$
  for $i \leftarrow 2, iterations$ **do**
    Sample θ_{prop} from proposal density g given θ_{old} and covariance matrix Σ
    Calculate acceptance probability $p_A(\theta_{\text{prop}}, \theta_{\text{old}}, data)$ given observed data
    Sample U from $\text{Uniform}(0, 1)$
    if $U < p_A$ **then**
      $\theta^{(i)} \leftarrow \theta_{\text{prop}}$
    else
      $\theta^{(i)} \leftarrow \theta_{\text{old}}$
    end if
    $\theta_{\text{old}} \leftarrow \theta^{(i)}$
  end for
  return $\theta^{(1)}, \ldots, \theta^{(iterations)}$
end procedure

In order to compute the likelihood conditioned on the initial infection count I_0 we assumed that I_t is only dependent on I_{t-1}:

$$\begin{aligned}
L(\theta) &= f(I|\theta) \\
&= f(I_n|I_{n-1}, \ldots, I_0, \theta) \cdot f(I_{n-1}|I_{n-2}, \ldots, I_0, \theta) \cdot \ldots \cdot f(I_1|I_0, \theta) \\
&= f(I_n|I_{n-1}, \theta) \cdot f(I_{n-1}|I_{n-2}, \theta) \cdot \ldots \cdot f(I_1|I_0, \theta) \\
&= \prod_{t=1}^{n} f(I_t|I_{t-1}, \theta) \\
&= \prod_{t=1}^{n} \frac{\exp(-\lambda_t)\lambda_t^{I_t}}{I_t!},
\end{aligned}$$

where $\lambda_t = \lambda_t(\theta, I_{t-1}) = \exp(\beta_0 + \alpha \log(I_{t-1} + 1))$. For easier computation we considered the logarithm:

$$\begin{aligned}
\log L(\theta) &= \sum_{t=1}^{n} \log\left(\frac{\exp(-\lambda_t)\lambda_t^{I_t}}{I_t!}\right) \\
&= \sum_{t=1}^{n} I_t \log(\lambda_t) - \log(I_t!) - \lambda_t.
\end{aligned}$$

As priors we assumed $\alpha \sim \mathrm{N}(0, \phi_\alpha), \beta_0 \sim \mathrm{N}(0, \phi_{\beta_0})$, independently. Thus,

$$\begin{aligned}
\pi_\alpha(x) &= \frac{1}{\sqrt{2\pi\phi_\alpha}} \exp\left(\frac{-x^2}{2\phi_\alpha}\right), \\
\pi_{\beta_0}(x) &= \frac{1}{\sqrt{2\pi\phi_{\beta_0}}} \exp\left(\frac{-x^2}{2\phi_{\beta_0}}\right), \\
\pi(\theta) &= \pi_\alpha(\alpha)\pi_{\beta_0}(\beta_0).
\end{aligned}$$

We set $\phi = (\phi_{\beta_0}, \phi_\alpha) = (100, 100)$. In a sensitivity analysis with ϕ being multiplied with $i \in \{0.01, 0.1, 1, 10\}$ we could not detect any relevant changes with respect to parameter estimates and acceptance rates. We also compared the results to the algorithm when using a flat prior, $\alpha \sim 1, \beta_0 \sim 1$ and did not find any relevant differences in the estimates.

In order to obtain a new θ candidate in the chain we utilized the proposal density

$$g(\theta_{\text{prop}}|\theta_{\text{old}}) = \mathrm{N}(\theta_{\text{old}}, C \cdot \Sigma),$$

where Σ and ϕ are given variances and C a constant matrix, which we assumed in the general case to be the identity matrix but enabled adaptation for a sensitivity analysis. For Σ we used the covariance matrix of the frequentist fit with the function `tsglm` from the `tscount` R package [123]. Furthermore, we also utilized its coefficient estimates perturbed by an additive value sampled from a Normal distribution with standard deviation 0.1 and mean 0 as starting values for the Bayesian methods.

In the following, let θ_{prop} be the proposal and θ_{old} the values of the previous draw in the MCMC. We sampled U from a $\mathrm{Uniform}(0, 1)$ distribution and accepted the proposal after comparing U with the factor

$$\begin{aligned}
\frac{f(\theta_{\text{prop}}|I)g(\theta_{\text{old}}|\theta_{\text{prop}})}{f(\theta_{\text{old}}|I)g(\theta_{\text{prop}}|\theta_{\text{old}})} &= \frac{f(I|\theta_{\text{prop}})\pi(\theta_{\text{prop}})g(\theta_{\text{old}}|\theta_{\text{prop}})}{f(I|\theta_{\text{old}})\pi(\theta_{\text{old}})g(\theta_{\text{prop}}|\theta_{\text{old}})} \\
&= \frac{L(\theta_{\text{prop}})\pi(\theta_{\text{prop}})g(\theta_{\text{old}}|\theta_{\text{prop}})}{L(\theta_{\text{old}})\pi(\theta_{\text{old}})g(\theta_{\text{prop}}|\theta_{\text{old}})}.
\end{aligned}$$

Since the proposal function is symmetric in the sense that $g(\theta_{\text{prop}}|\theta_{\text{old}}) = g(\theta_{\text{old}}|\theta_{\text{prop}})$, the factor for the acceptance probability p_A can be simplified. For numerical stability, we work on the log scale, and accept a proposal if

$$\log(U) \leq \log(p_A) = \min\left(\log(1), \log\left(\frac{L(\theta_{\text{prop}})\pi(\theta_{\text{prop}})}{L(\theta_{\text{old}})\pi(\theta_{\text{old}})}\right)\right).$$

If we consider

$$\begin{aligned}
\log(1) &= 0,\\
\log\left(\frac{L(\theta_{\text{prop}})}{L(\theta_{\text{old}})}\right) &= \sum_{t=1}^{n}\left(I_t\log\left(\lambda_t^{(\text{prop})}\right)-\log(I_t!)-\lambda_t^{(\text{prop})}\right)\\
&\quad-\sum_{t=1}^{n}\left(I_t\log\left(\lambda_t^{(\text{old})}\right)-\log(I_t!)-\lambda_t^{(\text{old})}\right)\\
&= \sum_{t=1}^{n}\left[I_t\left(\log\left(\lambda_t^{(\text{prop})}\right)-\log\left(\lambda_t^{(\text{old})}\right)\right)-\lambda_t^{(\text{prop})}+\lambda_t^{(\text{old})}\right],\\
\log\left(\frac{\pi(\theta_{\text{prop}})}{\pi(\theta_{\text{old}})}\right) &= \log\left(\frac{\pi_\alpha(\alpha^{(\text{prop})})\pi_{\beta_0}(\beta_0^{(\text{prop})})}{\pi_\alpha(\alpha^{(\text{old})})\pi_{\beta_0}(\beta_0^{(\text{old})})}\right)\\
&= \log\left(\pi_\alpha(\alpha^{(\text{prop})})\right)+\log\left(\pi_{\beta_0}(\beta_0^{(\text{old})})\right)\\
&\quad-\log\left(\pi_\alpha(\alpha^{(\text{old})})\right)-\log\left(\pi_{\beta_0}(\beta_0^{(\text{old})})\right)\\
&= -\frac{\alpha^{(\text{prop})^2}}{2\phi_\alpha}-\frac{\beta_0^{(\text{prop})^2}}{2\phi_{\beta_0}}+\frac{\alpha^{(\text{old})^2}}{2\phi_\alpha}+\frac{\beta_0^{(\text{old})^2}}{2\phi_{\beta_0}},
\end{aligned}$$

where $\lambda_t^{(\text{old})}=\lambda_t(\theta_{\text{old}},I_{t-1})$ and $\lambda_t^{(\text{prop})}=\lambda_t(\theta_{\text{prop}},I_{t-1})$, we obtain

$$\begin{aligned}
\log(p_A) = \min\Bigg(0,&-\frac{\alpha^{(\text{prop})^2}}{2\phi_\alpha}-\frac{\beta_0^{(\text{prop})^2}}{2\phi_{\beta_0}}+\frac{\alpha^{(\text{old})^2}}{2\phi_\alpha}+\frac{\beta_0^{(\text{old})^2}}{2\phi_{\beta_0}}\\
&+\sum_{t=1}^{n}\left[I_t\left(\log\left(\lambda_t^{(\text{prop})}\right)-\log\left(\lambda_t^{(\text{old})}\right)\right)-\lambda_t^{(\text{prop})}+\lambda_t^{(\text{old})}\right]\Bigg).
\end{aligned}$$

The acceptance probability is simplified, when using a reference prior, which eliminates the term before the summation symbol in the second factor. It can also be further computationally optimized by considering

$$\begin{aligned}
\sum_{t=1}^{n}I_t\log(\lambda_t) &= \sum_{t=1}^{n}I_t\beta_0+I_t\alpha\log(I_{t-1}+1) = \beta_0\sum_{t=1}^{n}I_t+\alpha\sum_{t=1}^{n}I_t\log(I_{t-1}+1)\\
&= \beta_0 sumi+\alpha sumli,\\
\sum_{t=1}^{n}\lambda_t &= \exp(\beta_0)\sum_{t=1}^{n}\exp(\alpha\log(I_{t-1}+1)) = \exp(\beta_0)\sum_{t=1}^{n}\exp(\log(I_{t-1}+1))^\alpha\\
&= \exp(\beta_0)\sum_{t=1}^{n}(I_{t-1}+1)^\alpha,
\end{aligned}$$

where $sumi=\sum_{t=1}^{n}I_t$ and $sumli=\sum_{t=1}^{n}I_t\log(I_{t-1}+1)$ depend only on the data and

can be calculated just once, before the sampler begins. Thus, the log of the acceptance probability can be written as:

$$\log(p_A) = \min\Bigg(0, -\frac{\alpha^{(\text{prop})^2}}{2\phi_\alpha} - \frac{\beta_0^{(\text{prop})^2}}{2\phi_{\beta_0}} + \frac{\alpha^{(\text{old})^2}}{2\phi_\alpha} + \frac{\beta_0^{(\text{old})^2}}{2\phi_{\beta_0}} \\ + sumi\left(\beta_0^{(\text{prop})} - \beta_0^{(\text{old})}\right) + sumli\left(\alpha^{(\text{prop})} - \alpha^{(\text{old})}\right) \\ - \exp\left(\beta_0^{(\text{prop})}\right)\sum_{t=1}^{n}(I_{t-1}+1)^{\alpha^{(\text{prop})}} + \exp\left(\beta_0^{(\text{old})}\right)\sum_{t=1}^{n}(I_{t-1}+1)^{\alpha^{(\text{old})}}\Bigg)$$

The code for the Metropolis sampler is shown in B.2.

As mentioned we utilized the frequentist approach of the function `tsglm` from the `tscount` package to find suitable starting values but also to compare the results to the Bayesian methods [123]. The `tsglm` function generally employs quasi conditional maximum likelihood estimation, which simplifies for Poisson models to an ordinary maximum likelihood estimator [123]. A more detailed description of this method specifically for Poisson autoregression is given in Fokianos *et al.* [124]. The `tsglm` function returned a mean estimate of α at 0.84 with a standard error of < 0.01 and a 95% confidence interval between 0.78 and 0.80. The mean estimate of β_0 was 1.14 with a standard error 0.02 and confidence interval limits at 1.10 and 1.19.

We compared multiple commonly used MCMC programs employing different algorithms and our Metropolis sampler. The Metropolis sampler was written and run in R while the others were accessed via R to enable post-processing of the output in similar fashion. WinBUGS as one of the first popular statistical software applications for Bayesian analysis employed Gibbs sampling, but was only usable on Windows platforms [125]. In order to have larger applicability, OpenBUGS and Just Another Gibbs Sampler (JAGS) [126] were developed. Although very similar, they were not identical. For example, OpenBUGS applied a different updating method [125]. While the source code of WinBUGS and OpenBUGS was written in Component Pascal, JAGS was written in C++, thereby increasing usability. JAGS also corrected some flaws present in WinBUGS at the time. Despite the disparities, WinBUGS, JAGS, and OpenBUGS all rely on BUGS scripts, which enabled us to use the same script in our implementation. Finally, we also incorporated the probabilistic programming language Stan. The main differences to the other methods are that Stan uses a more flexible and expressive programming language and that its MCMC technique is based on Hamiltonian Monte Carlo (HMC), which is more efficient and robust for complex posteriors than Gibbs or Metropolis-Hastings samplers [127]. The default sampler of Stan is the No-U-Turn Sampler, an adaptive variant of HMC sampling [128]. As it is based on HMC, it uses gradient information

in the log density function thereby limiting random walk behaviour and sensitivity to correlated parameters.

HMC was first described in Duane *et al.* [129], where it was still called Hybrid Monte Carlo. We will give a short introduction to the method and the No-U-Turn expansion and will therefore mostly rely on Hoffman & Gelman [128]. For a more detailed introduction to HMC, Neal *et al.* [130] explains the topic with a focus on one of the origins of the approach, molecular dynamics, specifically Hamiltonian dynamics, while Betancourt [131] describes the method in a detailed manner from a geometric point of view. HMC introduces for each model variable θ_d a momentum variable r_d. These are independently drawn from the standard normal distribution. Thus,

$$p(\theta, r) \propto \exp\Big(\mathcal{L}(\theta) - \frac{1}{2} r \cdot r\Big),$$

where $\mathcal{L}(\theta)$ is the logarithm of the joint density of θ and $x \cdot y$ denotes the inner product of the vectors x and y. Typically, a leapfrog integrator, simulating a Hamiltonian dynamics evolution, is used to propose a move to a new point. Thereby, the position and momentum vectors after step size ϵ are given by

$$\begin{aligned} r^{(t+\epsilon/2)} &= r^{(t)} + (\epsilon/2)\nabla_\theta \mathcal{L}(\theta^{(t)}) \\ \theta^{(t+\epsilon)} &= \theta^{(t)} + \epsilon r^{t+\epsilon/2} \\ r^{(t+\epsilon)} &= r^{(t+\epsilon/2)} + (\epsilon/2)\nabla_\theta \mathcal{L}(\theta^{(t+\epsilon)}), \end{aligned}$$

where ∇_θ denotes the gradient with respect to θ.

A standard sampling strategy is described in Algorithm 2, where M is the targeted number of samples. For each sample m the momentum variables are sampled from a standard normal distribution, analogous to Gibbs sampling. These momentum variables and the previous position variables $\theta^{(m-1)}$ are used as starting values for the leapfrog process with L updates and step size ϵ. The resulting momentum and position variables are used as proposals and are accepted according to the Metropolis algorithm.

The effectiveness of HMC depends strongly on the choice of appropriate values for ϵ and L. The No-U-Turn Sampler eliminates the need to set L manually and thereby removes a big pitfall of the approach. It does so by simulating a Hamiltonian trajectory by using the leapfrog integrator forwards and backwards until it starts making a U-turn and begins going back towards another position on the trajectory. Since retracing its steps will likely be inefficient the simulation is stopped and a point is sampled from the computed path. In order to find suitable ϵ, the dual averaging algorithm of Nesterov [132] can be utilized.

Algorithm 2 Hamiltonian Monte Carlo as described in Hoffman & Gelman [128]

procedure HAMILTONIANSAMPLER($startvalue, stepsize, updates, density, samplesize$)
  $\theta^{(0)} \leftarrow startvalue$
  $\epsilon \leftarrow stepsize$
  $L \leftarrow updates$
  $\mathcal{L} \leftarrow density$
  $M \leftarrow samplesize$
  for $m \leftarrow 1, M$ **do**
    Sample $r^{(0)} \sim \mathrm{N}(0, I)$
    $\theta^m \leftarrow \theta^{(m-1)}, \tilde{r} \leftarrow r^{(0)}$
    for $i \leftarrow 1, L$ **do**
      $\tilde{\theta}, \tilde{r} \leftarrow$ Leapfrog$(\tilde{\theta}, \tilde{r}, \epsilon)$
    end for
    With probability $\min\left(1, \frac{\exp\left(\mathcal{L}(\tilde{\theta} - \frac{1}{2}\tilde{r} \cdot \tilde{r})\right)}{\exp\left(\mathcal{L}(\theta^{(m-1)} - \frac{1}{2}r^{(0)} \cdot r^{(0)})\right)}\right)$, set $\theta^{(m)} \leftarrow \tilde{\theta}, r^{(m)} \leftarrow -\tilde{r}$
  end for
  return $\theta^{(1)}, \ldots, \theta^{(m)}$
end procedure

function LEAPFROG(θ, r, ϵ)
  $\tilde{r} \leftarrow r + (\epsilon/2)\nabla_\theta \mathcal{L}(\theta)$
  $\tilde{\theta} \leftarrow \theta + \epsilon\tilde{r}$
  $\tilde{r} \leftarrow \tilde{r} + (\epsilon/2)\nabla_\theta \mathcal{L}(\tilde{\theta})$
  return $\tilde{\theta}, \tilde{r}$
end function

We used the packages `R2WinBUGS` [133], `R2jags` [134], and `rstan` [135] as R interfaces to the respective software. Table 2.2 compares the results of these methods and the Metropolis sampler applied to daily infections in Munich.

There are numerous metrics for determining convergence of MCMC chains [136]. Probably the most widely used is $\widehat{R}$, first defined by Gelman & Rubin [137]. The basic idea of the metric is that for multiple MCMC chains, the variance of the chains together should be higher than the variance of the individual chains, if the simulations have not mixed well. Thus, for each scalar of interest the standard deviation from all chains together is divided by the root of the mean square of the individual within-chain standard deviations to obtain the potential scale reduction $\widehat{R}$. The definition evolved over the last years to correct shortcomings of the original definition. Gelman *et al.* [138] introduced the split-$\widehat{R}$ variant, which the computation of the `rstan` package is based on. We define it below with the notation of Vehtari *et al.* [139]. Let B and W be the between- and within-chain variances of the scalar θ with respect to M chains each with N draws. Then

	Method	Mean	SE	Quantile			$\widehat{R}$	Runtime
				2.5%	50%	97.5%		[seconds]
β_0	Metropolis R	1.15	0.02	1.11	1.15	1.20	1.00	4.7
β_0	WinBUGS	1.10	0.02	1.06	1.10	1.15	1.01	8.0
β_0	OpenBUGS	1.10	0.02	1.06	1.10	1.15	1.01	46.3
β_0	Jags	2.60	0.03	2.56	2.60	2.64	1.00	36.8
β_0	Stan	1.15	0.02	1.11	1.15	1.20	1.00	50.2
α	Metropolis R	0.79	0.00	0.78	0.79	0.80	1.00	4.7
α	WinBUGS	0.80	0.00	0.79	0.80	0.81	1.01	8.0
α	OpenBUGS	0.80	0.00	0.79	0.80	0.81	1.01	46.3
α	Jags	0.54	0.00	0.54	0.54	0.55	1.00	36.8
α	Stan	0.79	0.00	0.78	0.79	0.80	1.00	50.2

Table 2.2: Results for β_0 and α across Bayesian sampling methods with respect to daily COVID-19 infection numbers in Munich from 2020-02-24 to 2021-03-16. SE is the standard error, $\widehat{R}$ is an indicator for convergence with values < 1.01 indicating convergence. For the Metropolis sampler the function `Rhat` from the package `rstan` was used to calcuate $\widehat{R}$, the other methods reported it as standard output of their function. In the case of WinBUGS the runtime only comprises the access time from R to the program, since it requires user input.

B and W are defined as

$$B = \frac{N}{M-1}\sum_{m=1}^{M}(\bar{\theta}^{(\cdot m)} - \bar{\theta}^{(\cdot\cdot)})^2, \qquad \bar{\theta}^{(\cdot m)} = \frac{1}{N}\sum_{n=1}^{N}\theta^{(nm)}, \bar{\theta}^{(\cdot\cdot)} = \frac{1}{M}\sum_{m=1}^{M}\bar{\theta}^{(\cdot m)},$$
$$W = \frac{1}{M}\sum_{m=1}^{M}s_m^2, \qquad s_m^2 = \frac{1}{N-1}\sum_{n=1}^{N}(\theta^{(nm)} - \bar{\theta}^{(\cdot m)}),$$

with $\theta^{(nm)}$ the nth draw in chain m, $\bar{\theta}^{(\cdot m)}$ the average of all draws from chain m, $\bar{\theta}^{(\cdot\cdot)}$ the average of all draws and s_m^2 the variance of the mth chain. The marginal posterior variance $\mathrm{var}(\theta|y)$ is then estimated by a weighted average of the above variances:

$$\widehat{\mathrm{var}}^{+}(\theta|y) = \frac{N-1}{N}W + \frac{1}{N}B.$$

From this follows the definition

$$\widehat{R} = \sqrt{\frac{\widehat{\text{var}}^{+}(\theta|y)}{W}},$$

which declines to 1 as $N \to \infty$ for an ergodic process. It is called split-$\widehat{R}$ because the chains are split in half beforehand such that M is twice the number of simulated chains. Vehtari *et al.* [139] suggest at least four chains and $\widehat{R} < 1.01$ to confirm convergence. Therefore, we sampled four chains each with 15,000 iterations and discarded a burn-in of 5,000 draws.

Table 2.2 shows that the MCMC estimates are almost identical and $\widehat{R} < 1.01$ for all methods except for JAGS. The posterior estimate $\alpha = 0.79$ is equal to the estimate of the frequentist method, `tsglm`, both with a standard error < 0.01. The posterior estimate $\beta_0 = 1.14$ and its standard error are also equal to the `tsglm` results. The Stan algorithm was by far the slowest, while the other implementations were still considerably slower than the manually coded Metropolis sampler for a given number of iterations. However, as can be seen in Figure 2.7, Stan converges much faster than the Metropolis sampler. This illustrates that Stan requires fewer iterations and a shorter burn-in.

In order to confirm that we obtain sensible results from our methods we simulated data similar to the available data with a broad range of parameter combinations, fit the data with a selection of the above methods and compared the estimates with the actual values. We iteratively sampled 500 observations from the distribution described in Equation 2.2 for each parameter combination. For the coefficient estimates, we used `tsglm`, the Metropolis sampler, and Stan on this simulated data. As before we used $\mathrm{N}(0, \phi_\alpha)$ as independent priors for the two coefficients and slightly disturbed coefficient estimates of `tsglm` as starting values for the Bayesian methods. In order to reduce computational time Metropolis only samples 3,000 times including a burn-in of 1,000. For Stan half of the iterations and burn-in length were used. The simulation and fitting were repeated 500 times. The whole procedure was executed for 16 combinations of $\beta_0 \in \{0, 0.5, 1, 2\}$ and $\alpha \in \{0, 0.3, 0.6, 0.9\}$ model values. For illustration, we plotted a histogram for one simulation of each of the combinations (Figure B.1 in the appendix).

In the following, we will only consider Bayesian results if the chains of the sampling converged. We assume convergence if $\widehat{R} < 1.01$. The comparison of the biases, which are the estimated value minus the actual parameter value, shows that the estimates are overall close to the actual value and there is mostly no systematic deviation. The differences only become apparent for greater values of the model parameters (Figure B.2 in the appendix). Using `tsglm` underestimates β_0 and overestimates α in these

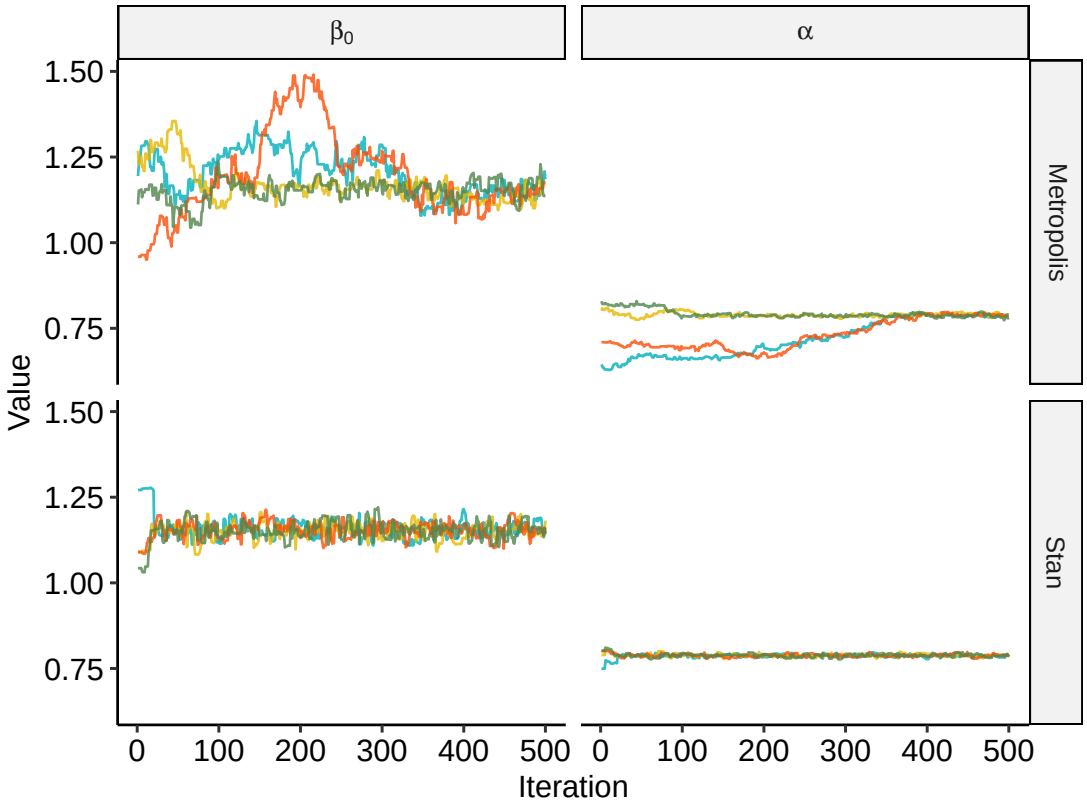

Figure 2.7: Trace plots of Metropolis sampler and Stan fitting Poisson model of COVID-19 infections in Munich. Only the first 500 draws are shown, which are discarded as burn-in.

cases. However, in two of these three parameter combinations, the models generate such high values that they would not be relevant for our use case (see Figure B.1). Nonetheless, in the model with $\beta_0 = 0.5$ and $\alpha = 0.9$, which might simulate values in a domain interesting for us, this tendency of estimate deviation is already slightly noticeable. The Bayesian methods are influenced by this since their starting values are based on `tsglm`. This leads to non-convergence, which could be improved by more sampling steps, otherwise, Stan and Metropolis have comparably good estimation results and seem to be more reliable for the more extreme cases.

For 95% posterior credible intervals we use the 2.5% and 97.5% quantiles of the posterior samples for the respective parameters. We defined coverage rates as the fraction of times the initial parameters fell inside their credible intervals or the confidence interval in the case of `tsglm` of the 500 simulated datasets. Figure 2.8 shows the coverage rates of the confidence intervals and the credible intervals. Although `tsglm`'s estimate is usually very close, its 95% confidence interval is too small to cover the model

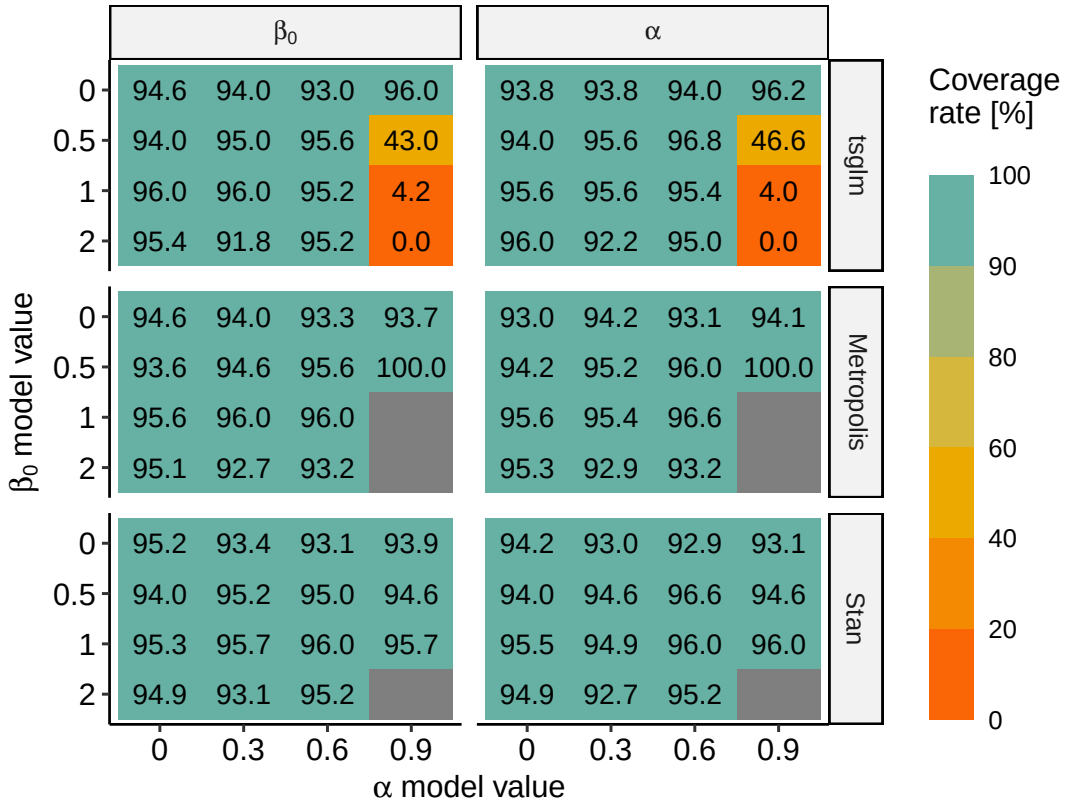

Figure 2.8: Coverage rates in percent of the confidence interval in the case of `tsglm` and credible intervals in the case of Metropolis and Stan for simulated data. One simulation comprises 500 observations distributed by the autocorrelated Poisson model defined in Equation 2.2 with different combinations of β_0 and α values. 500 simulations were performed for each combination. Gray cells indicate non-convergence for all simulations.

parameter if α is very large, while the Bayesian methods are again more dependable for the more extreme conditions. However, in general, all three models show coverage rates close to 95% suggesting intervals neither too large nor too small.

2.3.2 Negative Binomial Model

A Poisson distribution has the limitation that the variance is equal to the expectation. Negative binomial distributions can be considered a generalization with an overdispersion parameter allowing the variance to exceed the expected value. In the following, we show a suitable parametrization in terms of the negative binomial distribution and derive important properties. We assumed that $I_t \sim \mathrm{NegBin}(\lambda_t, \psi)$, such that λ_t is the mean of

I_t and $\lambda_t(1+\psi\lambda_t)$ the variance. This can be represented as:

$$\begin{aligned} I_t|\mu_t &\sim \text{Poisson}(\mu_t) \quad \mu_t > 0 \quad \forall t \in \{1,\dots,n\} \\ \mu_t|\lambda_t,\psi &\sim \text{Gamma}(\psi^{-1}, \frac{\psi^{-1}}{\lambda_t}) \\ \log(\lambda_t(\alpha,\beta_0)) &= \beta_0 + \alpha\log(I_{t-1}+1) \\ \psi &\geq 0 \\ \alpha,\beta_0 &\in \mathbb{R} \end{aligned}$$

We show that the negative binomial is equivalent to the above Poisson-Gamma mixture with a parametrization with respect to the mean λ and the overdispersion parameter $r = \psi^{-1}$. For easier reading we skip the t index for λ and μ in the following equation:

$$\begin{aligned} \int_0^\infty f_{\text{Poisson}}(y;\mu) f_{\text{Gamma}}(\mu;r,\frac{r}{\lambda})\,d\mu &= \int_0^\infty \frac{\mu}{y!}\exp(-\mu)\frac{\mu^{r-1}}{\Gamma(r)}\left(\frac{r}{\lambda}\right)^r \exp\left(-\mu\frac{r}{\lambda}\right) d\mu \\ &= \frac{r^r}{\lambda^r y!\Gamma(r)}\int_0^\infty \mu^{y+r-1}\exp\left(-\mu\frac{\lambda+r}{\lambda}\right) d\mu \\ &= \frac{r^r}{\lambda^r y!\Gamma(r)}\int_0^\infty \left(\nu\frac{\lambda}{\lambda+r}\right)^{y+r-1}\exp(-\nu)\frac{\lambda}{\lambda+r}\,d\nu \\ &= \frac{r^r}{\lambda^r y!\Gamma(r)}\left(\frac{\lambda}{\lambda+r}\right)^{y+r}\Gamma(y+r) \\ &= \frac{\Gamma(y+r)}{y!\Gamma(r)}\left(\frac{\lambda}{\lambda+r}\right)^y\left(\frac{r}{\lambda+r}\right)^r \\ &= \frac{\Gamma(y+r)}{y!\Gamma(r)}\left(\frac{\lambda}{\lambda+r}\right)^y\left(1-\frac{\lambda}{\lambda+r}\right)^r \\ &= f_{\text{NegativeBinomial}}(y;\lambda,r), \end{aligned}$$

where we applied the integral transformation $\nu = \mu(1+r/\lambda)$. In the literature the distribution is often defined by another parametrization $p = \frac{\lambda}{r+\lambda}$.

We used the shorthand $\lambda_t = \lambda_t(\theta, I_{t-1}) = \exp(\beta_0 + \alpha\log(I_{t-1}+1))$ in order to com-

pute the likelihood:

$$\begin{aligned}
L(\theta) &= P(I, \mu|\theta) \\
&= P(I_n, \ldots, I_0|\theta) \\
&= P(I_n|I_{n-1}, \ldots, I_0, \theta) \cdot \ldots \cdot P(I_1|I_0, \theta) \\
&= P(I_n|I_{n-1}, \theta) \cdot \ldots \cdot P(I_1|I_0, \theta) \\
&= \prod_{t=1}^{n} P(I_t|I_{t-1}, \theta) \\
&= \prod_{t=1}^{n} \frac{\Gamma(I_t + r)}{I_t!\Gamma(r)} \left(\frac{\lambda_t}{\lambda_t + r}\right)^{I_t} \left(1 - \frac{\lambda_t}{\lambda_t + r}\right)^{r} \\
&= \Gamma(r)^{-n} \prod_{t=1}^{n} \frac{\Gamma(I_t + r)}{I_t!} \left(\frac{\lambda_t}{\lambda_t + r}\right)^{I_t} \left(\frac{r}{\lambda_t + r}\right)^{r}
\end{aligned}$$

The log-likelihood can then be written as:

$$\begin{aligned}
\log L(\theta) = &-n \log\left(\Gamma(r)\right) + \sum_{t=1}^{n} \big(\log(\Gamma(I_t + r)) - \log(I_t!) \\
&+ I_t(\log(\lambda_t)) - I_t(\log(\lambda_t + r)) + r(\log(r)) - r(\log(\lambda_t + r))\big) \\
= &\, n\big(r(\log(r)) - \log\left(\Gamma(r)\right)\big) + \sum_{t=1}^{n} \big(\log(\Gamma(I_t + r)) - \log(I_t!) \\
&+ I_t(\log(\lambda_t)) - I_t(\log(\lambda_t + r)) - r(\log(\lambda_t + r))\big)
\end{aligned}$$

Inspired by Gelman *et al.* [140] we suggest $\alpha \sim \mathrm{Cauchy}(0, \phi_\alpha), \beta_0 \sim \mathrm{Cauchy}(0, \phi_{\beta_0})$ as suitable priors. Thus with the center $\mu = 0$ we define

$$\begin{aligned}
\pi_\alpha(x) &= \frac{\phi_\alpha^2}{\pi\phi_\alpha((x - \mu)^2 + \phi_\alpha^2)} = \frac{\phi_\alpha^2}{\pi\phi_\alpha(x^2 + \phi_\alpha^2)} \\
\pi_{\beta_0}(x) &= \frac{\phi_{\beta_0}^2}{\pi\phi_{\beta_0}((x - \mu)^2 + \phi_{\beta_0}^2)} = \frac{\phi_{\beta_0}^2}{\pi\phi_{\beta_0}(x^2 + \phi_{\beta_0}^2)}
\end{aligned}$$

with $\phi_\alpha = 2.5$ and $\phi_{\beta_0} = 10$. Similar to the Poisson we can employ for the proposal functions:

$$g(\theta_{\mathrm{prop}}|\theta_{\mathrm{old}}) = \mathrm{N}(\theta_{\mathrm{old}}, C \cdot \Sigma).$$

As defined above, the variance of the negative binomial distribution is $\lambda_t(1 + \psi\lambda_t)$.

Therefore, we set

$$\psi_{\text{prop}} = \frac{1}{n}\sum_{t=1}^{n}\frac{\sigma^2 - \lambda_t((\alpha,\beta_0)_{\text{prop}})}{\lambda_t((\alpha,\beta_0)_{\text{prop}})^2},$$

where σ^2 is the variance estimate of the response data.

For our later analyses, we also wanted to be able to incorporate various covariates and further autoregressive components. Thus we define a more general negative binomial model parametrization similar to Liboschik *et al.* [123], where we consider the history of the process until $t-1$, $\mathcal{F}_{t-1}$, for the infection count I_t:

$$\begin{aligned} I_t|\mathcal{F}_{t-1} \sim & NegBin(\lambda_t, \psi) \\ \log(\lambda_t) = & \beta_0 + \\ & \beta_1 \log(I_{t-i_1}+1) + \ldots + \beta_p \log(I_{t-i_p}+1) + \\ & \alpha_1 \log(\lambda_{t-j_1}) + \ldots + \alpha_q \log(\lambda_{t-j_q}) + \\ & \eta_1 X_{t,1} + \ldots + \eta_k X_{t,k}, \end{aligned} \tag{2.3}$$

where $X_{t,v}$ is the value of covariate v at time t, k is the number of included covariates, and $i, j \in 1, \ldots, t-1$.

Regarding the fitted models, we focused on the exponentiated coefficient estimates of Poisson and negative binomial models, which are called incidence rate ratios. Epidemiologically, the incidence rate ratio is defined as the number of new cases of a disease during a given time divided by the corresponding person-time of the observation [141]. We consider that the incidence is the number of COVID-19 infections divided by the population p of an area a, which we assume stays constant during our analysis time for each district, for any fixed time t. We divide the incidence by the length s of the time period of t and denote it $IR_t^{(a)}$. Let us consider the expected incidence rates of two areas, a and b, that differ in only one covariate v by one unit at time $t-1$. Then, we get for the ratio of the incidence rates at t considering Equation 2.3:

$$\begin{aligned}
\frac{E\left(IR_t^{(a)}\right)}{E\left(IR_t^{(b)}\right)} &= \frac{\frac{1}{ps}E\left(I_t^{(a)}\right)}{\frac{1}{ps}E\left(I_t^{(b)}\right)} \\
&= \frac{\lambda_t^{(a)}}{\lambda_t^{(b)}} \\
&= \frac{\exp\left(\beta_0 + \ldots + \eta_v(X_{t,v}+1) + \ldots + \eta_k(X_{t,k})\right)}{\exp\left(\beta_0 + \ldots + \eta_v X_{t,v} + \ldots + \eta_k(X_{t,k})\right)} \\
&= \frac{\exp\left(\eta_v(X_{t,v}+1)\right)}{\exp\left(\eta_v X_{t,v}\right)} \\
&= \exp\left(\eta_v\right),
\end{aligned}$$

where $\lambda_t^{(a)}, \lambda_t^{(b)}$ and $I_t^{(a)}, I_t^{(b)}$ are the means and infections of the areas a and b, respectively. Considering the final equation, we see that the result is independent of t. Thus, similar to the Cox proportional hazards model, the vth exponentiated coefficient at any time is the ratio of the incidence rates with a unit increase in covariate v assuming that all other covariates stay the same.

2.3.3 Selected model metrics

Variable selection in regression is a common challenge. Information criteria are useful tools in data-driven approaches to select variables. One of the most widely used is the Akaike Information Criterion (AIC) [142], which is defined by

$$AIC = -2\log\left(\widehat{L}\right) + 2k,$$

where $\widehat{L}$ is the maximum likelihood and k is the number of estimated parameters in the model. The BIC [65] is an alternative to the AIC, as already mentioned in section 2.1.3, and penalizes the number of parameters more severely. It does so by including the logarithm of the sample size n as penalizing factor:

$$BIC = -2\log\left(\widehat{L}\right) + \log(n)k.$$

The quasi-likelihood under the independence model criterion (QIC) expands the AIC in the case of regression analyses based on generalized estimating equations (GEE) by employing the quasi-likelihood instead of the likelihood and adjusting the penalty term

[143]. Thus, the QIC is defined as

$$QIC = -2Q + 2\text{trace}(\hat{\Omega}_I \hat{V}_r),$$

where Q is the quasi-log-likelihood, $\hat{\Omega}_I$ can be considered an empirical quasi-log-likelihood information matrix [144], and $\hat{V}_r$ is a robust covariance estimator.

To measure the accuracy of the prediction models, we used the Symmetric Median Absolute Percentage Error (SMDAPE). It is defined by the median of the differences between prediction, P_t, and actual observation, O_t, divided by the mean of the two values for all time points t [145]:

$$\text{SMDAPE} = \text{median}\left(\frac{|P_t - O_t|}{(P_t + O_t)/2}\right).$$

There are various other measures, which are suitable for count predictions. We chose SMDAPE since we did not have values close to zero, which may have led to instability, it is intuitive, and it adjusts weights by the respective scale. For example, a forecast error of 100 cases was by far not as concerning if the actual observation was at 2000 as it would be if the observed value was 200.

Chapter 3

Results

The following sections discuss results of four studies completed as part of this thesis. The first three are already published or submitted for publication, with the primary statistical methods explained in Chapter 2 and the background, data, results, limitations, and conclusions appearing in the respective publications. The fourth topic concerning COVID-19 longitudinal analyses, which is not yet submitted for publication, is discussed in detail.

The first study analysed survival outcomes following kidney transplantations in European countries and focused on the effect of a single risk factor, the allocation type. The goal was to determine if there was a significant association with transplant outcomes, including overall survival, death with functioning graft, and graft loss, which could substantially influence allocation of grafts. The second study created an online risk prediction tool to incorporate the most important predictors of kidney transplant outcomes, estimate the corresponding risks, and thus allow immediate quantitative-based decisions on graft allocations in clinical practice. The third and fourth studies shifted from patient-level survival data in kidney transplants to population-level counts with respect to COVID-19 over time. The third study focused again on a subset of risk factors to determine the significance of their association with the outcome. In this case, the outcome was the number of cumulative deaths due to COVID-19 for individual districts in Germany in the two years after the start of the pandemic. Risk factors of particular interest were air pollution measures. The cross-sectional approach analysed time-aggregated data over different spatial entities distinguishing it from longitudinal analyses on COVID-19 infection time series within single spatial entities such as in the fourth study. In this final study, we modelled COVID-19 infections on a weekly basis for selected German districts, thereby evaluating various time-dynamic risk factors as covariates, estimating their association with infection counts, and determining the potential of one-week-ahead predictions.

3.1 Kidney transplantation after rescue allocations

*Assfalg, V., **Miller, G.**, Stocker, F., et al. (2021). Kidney Transplantation After Rescue Allocation—the Eurotransplant Experience: A Retrospective Multicenter Outcome Analysis. Transplantation 106, 1215–1226 (2022).*

The Eurotransplant (ET) consortium provides two major waiting list programs for potential kidney transplant recipients, the Eurotransplant Kidney Allocation System and the Eurotransplant Senior Program, the latter limited to donors and recipients who are at least 65 years old. These define the standards for distribution of available organs, incorporating metrics such as time on waiting list, immunological matching, or regional distance. In the event that kidney transplants fail to be allocated through the standard system due to impending organ loss, the process converts to a rescue allocation. In such cases, allocation processes can deviate from previously required metrics in order to avoid organ loss.

We analysed deceased donor kidney transplantations of the ET program between 2006 and 2018 in order to determine if rescue allocations had significantly worse outcomes in comparison to standard allocations. The dataset contained 2,422 rescue and 25,481 standard allocated kidneys. We considered three different outcomes after the kidney transplant, patient survival, death with functioning graft, and graft survival. We applied a Cox proportional hazards model to analyse patient survival and the Fine-Gray proportional regression model for the competing risks, death with functioning graft and graft survival. We considered known risk factors that were routinely collected by ET. In addition to rescue versus standard allocation, risk factors included recipient gender, immunological matching, recipient and donor age, cold ischemia time, time on waiting list, indication whether the recipient had diabetes, and number of kidney transplants. Cold ischemia time was defined as the period of cold organ preservation. The senior allocation system had different patient characteristics and processes compared to the general allocation system, including higher recipient and donor age as well as shorter cold ischemia and waiting time. Therefore, we performed the analysis separately for the respective datasets.

Rescue transplants were associated with worse outcomes in univariate analyses. When accounting for other risk factors, rescue allocations were associated with longer survival for all three outcomes, although not significantly in most cases. On the other hand, poor immunological matching, higher donor and recipient age, increased cold ischemia time, longer waiting time, diabetes, and re-transplants were all associated with worse outcomes. Conclusions show that facing donor shortages coupled with

long waiting times, rescue allocations may provide a beneficial option to avoid loss of available organs.

Author contribution

Gregor Miller processed the data, performed all statistical analyses, and wrote the manuscript. Volker Aßfalg, Donna Ankerst, and Norbert Hüser participated in writing the manuscript. All authors participated in the manuscript revision.

3.2 A risk prediction tool for kidney transplants

***Miller, G.**, Ankerst, D. P., Kattan, M. W., et al. (2022). Kidney Transplantation Outcome Predictions (KTOP): A Risk Prediction Tool for Kidney Transplants from Brain-dead Deceased Donors Based on a Large European Cohort. European Urology. ISSN: 0302-2838 (2022).*

Relationships between risk factors and success of kidney transplantation have been well studied. For the United States in particular, online tools have already been developed to calculate corresponding risks. Allocation systems in the US are limited to a certain number of risk factors and are not directly transferable to the European population for which no online tool exists. The availability of a European-tailored risk tool is crucial for impacting clinical practice and decision-making as an estimate of transplant performance determines whether a transplantation should be performed with a given available organ. Risk estimates provide more confidence than anecdotal medicine, where individual clinicians recall specific experiences that may be subject to error. Furthermore, they can help manage patient expectations, when a graft is not yet available. Therefore, the aim of this study was to develop the first European online risk tool for kidney transplantation outcomes.

The survival analyses were based on the same kidney transplant dataset as for the first project, using patient death, graft loss, and death with functioning graft as outcomes. We distinguished between recipient-only and recipient-donor risk factors, leading to six models overall, flexibly allowing risk predictions when all recipient and donor characteristics were known, but also when donor information was not yet available. For the selection of variables to be included in each of the models we used backward variable selection employing the BIC while continuously adapting the dataset to the incorporated risk factors and the corresponding complete cases. As alternative, we also applied the imputation method MICE on the dataset, however there was no increase in

model accuracy. Recipient age, recipient sex, re-transplantation, number of transplants, cystic disease, diabetes, time on waiting list, and allocation program were in the recipient-only risk factor set. The recipient-donor set also included donor age, donor sex, cold ischemia time, allocation process, and immunological matching.

We re-trained models on a subset of the data and performed calibration and decision curve analyses on the remaining validation data. With 60%, Germany had the highest proportion of transplants in the ET dataset. There were also large differences for some risk factors between the countries. In order to determine if the models performed significantly different for some countries, we considered four country groups, Austria, Belgium, Germany, and the other countries. We selected one country group to validate the AUC, for five and ten year survival, and the coefficient estimates of the model that was fit on data of the remaining countries. This was iterated for all country groups to determine if there were any relevant differences.

Diabetes, re-transplantation, and recipient age were associated worse overall survival and survival with functioning graft. Diabetes, older donor age, and long cold ischemia time were associated with an increase in the graft loss hazard. The developed models achieved ten-year AUC values of up to 81%. Risks were well-calibrated and decision curves showed net benefit. Comparisons between countries showed only small deviations in effect estimation and accuracy.

The models were incorporated in an online risk tool at `riskcalc.org/ktop` (Figure 3.1) in order to support physicians and patients facing kidney transplants and potential organ donations. This reproducible and consistent tool fills a previously vacant role estimating the progress after such a transplantation both before and after a donor has been identified, particularly in a European context. We hope that the online tool fosters independent external validation. Having demonstrated its usefulness, the tool should be a step toward further development of transplant risk calculators that facilitate decisions about kidney and other organ allocation. Considering the importance of bench-to-bedside practice, public health, and ecological efforts to advance scientific discoveries and translate them into daily applications, efforts such as online risk calculators for accessible results will evolve even further in the future for example with a spatial distinction or continuous risk assessment after transplantation [146].

Author contribution

Gregor Miller analysed and interpreted the data, conducted all statistical analyses, designed and created the online tool, and drafted the manuscript. Gregor Miller, Donna Ankerst, and Volker Aßfalg were responsible for the study concept. Volker Aßfalg

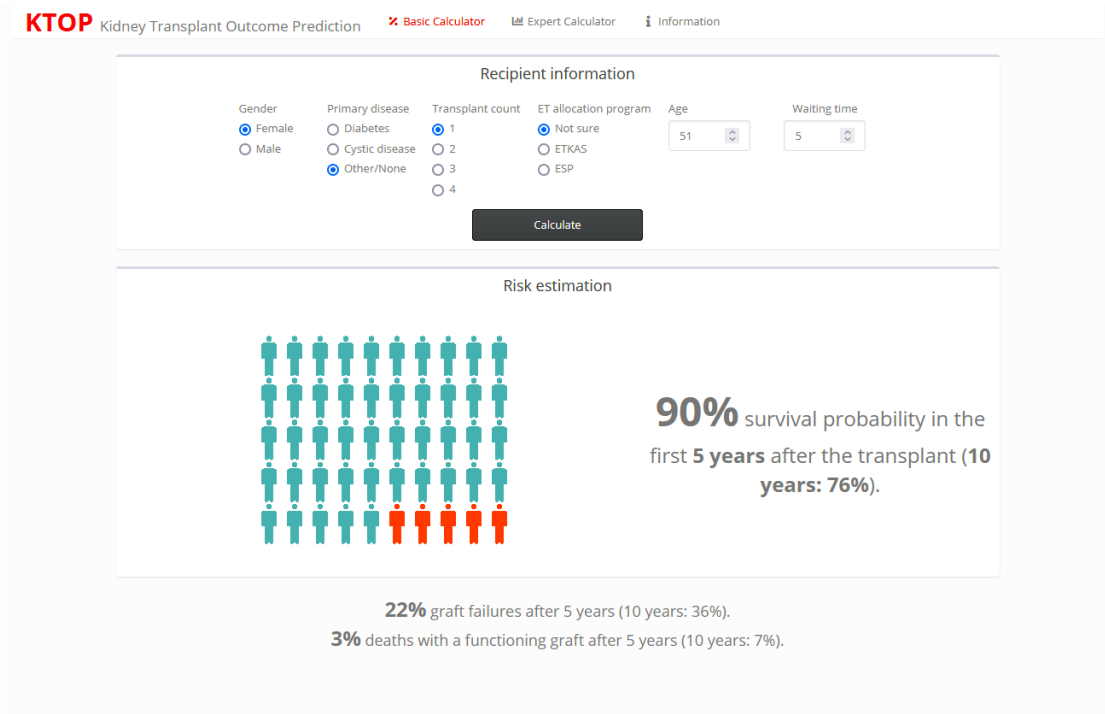

Figure 3.1: Screenshot of the online risk calculator for kidney transplants KTOP available at `riskcalc.org/ktop`.

participated in drafting the manuscript. All authors participated in the manuscript revision.

3.3 Associations of air pollution and COVID-19 in a cross-sectional analysis

***Miller, G.**, Menzel, A., Ankerst, D. P. (2022). Association between short-term exposure to air pollution and COVID-19 mortality in all German districts: The importance of confounders. In submission.*

Multiple studies have investigated the association between air pollution and the COVID-19 spreading process. Earlier studies provided preliminary results following the first few weeks after the pandemic outbreak and typically accounted for only few potential if any confounding factors, making results highly heterogeneous.

In this study, we sought to examine the association between short-term air pollution and COVID-19 mortality in Germany over two years, focusing particularly on sensitivity to choice of confounders. All 400 German districts were included and associations were

estimated in a cross-sectional analysis, where the number of deaths were cumulated over the analysis period for each district and modelled via negative binomial regression. We confined cumulative COVID-19 mortality analyses to two time periods, March 2020-February 2021 and March 2021-February 2022. We considered six air pollution variables, $PM_{2.5}$, PM_{10}, CO, NO, NO_2, and O_3, which were averaged over the time periods. To avoid multicollinearity among pollutants, we created separate models for each of the pollution variables with cumulative COVID-19 deaths for each of the two time periods. We identified and applied leading variable selection methods to account for 37 potential confounders, including change-in-estimate and information criteria methods. We then compared the selection approaches with models incorporating no and all confounders, while ensuring that the pollution variables were always included in the models.

Univariate analyses found significant positive associations with COVID-19, at least for the first time period, for O_3 and $PM_{2.5}$. CO, NO, and NO_2 showed negative associations. However, the inclusion of confounders dismissed these associations in the multivariate models for almost all selection approaches. The confounders most often selected were mobility, political voting behaviour, and age.

In conclusion, this study could not determine significant associations between COVID-19 mortality and air pollution, in contrast to a number of other studies. Assuming that there is an association, one possible reason for the lack of detection could be that the pollution levels were simply not high enough to have a relevant effect or that a cross-sectional study did not provide the sensitivity necessary to detect associations. However, the loss of significance in the multivariate models compared with univariate analysis, as well as a number of prior published studies, could indicate that air pollution served merely as a surrogate for confounding factors, such as mobility. The results from this study highlight the importance of appropriate confounder integration on the basis of initial exploratory analyses and sufficient data accumulation. Even though the predominant selection methods led to similar results, more significant effects could have generated more heterogeneous results across different choices of confounder variables for conclusion in the model. In such a case, especially when the focus is on inference, comparative methodologies need to be considered in order to draw reliable conclusions. This study showed how to implement the integration of multiple variable selection approaches, compare results, and draw conclusions in such a context.

Author contribution

Gregor Miller analysed and interpreted the data, conducted all statistical analyses, and drafted the manuscript. Gregor Miller and Donna Ankerst were responsible for the study

concept. All authors participated in the manuscript revision.

3.4 Longitudinal analyses of COVID-19 in Germany

In the final study of the thesis, we implemented longitudinal count models in order to investigate the association of weather variables with COVID-19 infections. The analysis was based on weekly data from four German cities, Berlin, Frankfurt, Hamburg, and Munich, between February 24, 2020, and March 6, 2022. We determined time series analysis as the most suitable approach in order to capture dynamic changes due to seasons and other confounders, such as mobility.

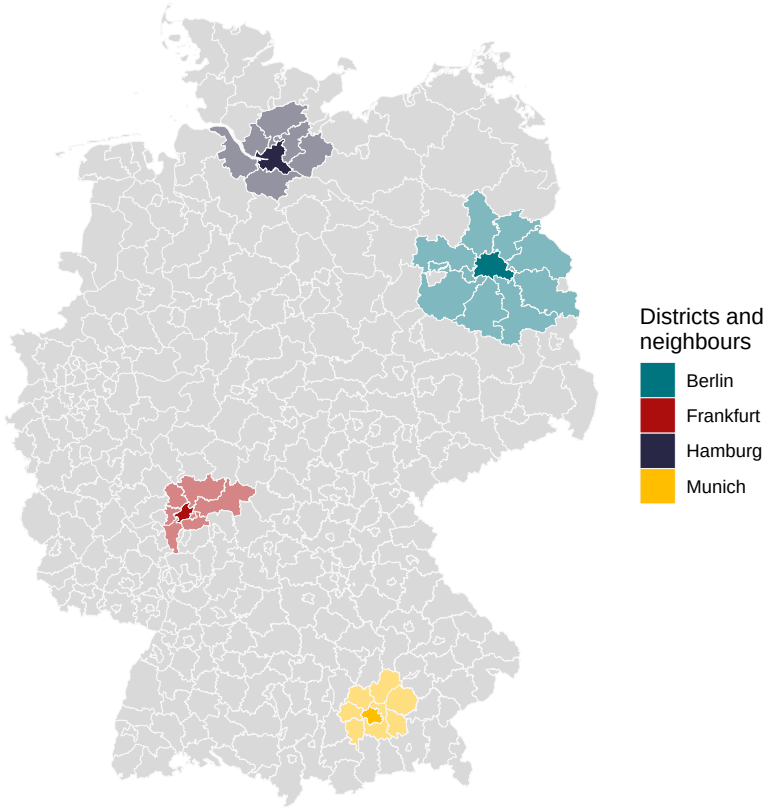

Figure 3.2: The four German districts for which longitudinal analyses were performed for the time period between February 24, 2020, and March 6, 2022, and their neighbouring districts.

3.4.1 Data

We assembled predictors and outcomes for the German administrative districts from a number of sources. The Federal Statistical Office provided population and area statistics for the latest reporting date on December 31, 2018 [147, 148]. We acquired the daily number of vaccinations from the RKI [149] as well as the daily number of COVID-19 infections [150] (dl-de/by-2-0 [151]). The daily vaccination rate was defined as the number of people who acquired full vaccination status as defined by RKI [149] on the specified day in the vaccination district divided by the population of that district. If the number of vaccinations administered in a district on a given day was less than five, they were not reported for that day but were added to the next day. We selected four of the 400 districts in Germany, each representing a major city, to perform the time series analysis. The four cities were Berlin, Frankfurt, Hamburg, and Munich, which have 3.7, 0.8, 1.8, and 1.5 million inhabitants and are situated in the north-east, north, west, and south of Germany, respectively (Figure 3.2). For analyses of the selected districts, neighbouring districts were also incorporated by calculating the log of their weekly sum of infections per 100,000 inhabitants. Before dividing infection counts by population counts, 1 was added to avoid having to calculate the log of 0. Data from February 24, 2020, to March 6, 2022, representing 106 weeks, was used for the analysis.

We extracted daily mobility data from the Google Community Mobility Report [152], which was only available on a state level for the 16 states in Germany. The mobility data quantified the change in the number of visits and length of stay at specific locations, including grocery stores, pharmacies, parks, residences, retail and recreational facilities, transit stations, and workplaces, compared to a reference period between January 3 and February 6, 2020. We averaged the daily values for each week.

The dates of country-wide lockdowns were extracted from government announcements [153–157]. The first lockdown started on March 22, 2020, and ended on May 6, 2020, when restrictions were at least partially lifted. A soft lockdown was implemented on November 2, 2020. A more severe lockdown with enhanced restrictions was then introduced on December 16, 2020, and ended on January 10, 2021. Season variables were defined so that winter refers to December, January, February; spring refers to March, April, May; summer refers to June, July, August; and autumn refers to September, October, November. In order to summarize daily indicators of season and lockdown to weekly indicators, we selected the value of the majority of days in the respective week.

We joined these statistics with geographic information on the administrative boundaries of the individual districts. The required geographic data of the districts was downloaded in `geojson` format with a simplification factor of 20 provided by Stabiszewski

& Ebert [158] (Geodatenzentrum © GeoBasis-DE / BKG 2018 (VG250 31.12., Data changed)). The original geographic information was adapted for the analyses as necessary, such that polygons only covering sea were excluded and the subdivision of the district Berlin was united.

Daily weather data was continuously queried and compiled from the open data resource of the German weather service [159], which provided daily measurements of several variables from 581 weather stations all over Germany. Six of the stations, which had an altitude above 1,000 metres, were excluded. We utilized inverse distance weighting to interpolate the station values for each district based on the distance between weather stations and the centroid of that district. A maximum of five stations were used to infer the value for each district for every day in the relevant time frame. We employed the function `idw` of the R package `gstat` [160, 161] to apply the interpolation. Figure 3.3 shows the locations of weather stations and an example interpolation on a specific day. We considered the daily mean wind speed in meters per second, precipitation in millimetres, sunshine duration in hours, mean temperature in degrees Celsius, and mean relative humidity in percent. Additionally, we calculated daily absolute humidity AH in g/m^3 for each district based on the temperature T and relative humidity RH [162] as:

$$AH = \frac{6.112 \times \exp\left(\frac{17.67 \times T}{T+243.5}\right) \times 2.1674 \times RH}{273.15 + T}.$$

3.4.2 Methods

Since we have previously noted the importance of confounders for this data, particularly mobility, the inclusion of other risk factors was necessary. This suggested the use of the frequentist R package `tscount`, due to the high computational cost of Bayesian models with increased model complexity and the similarity of the results shown previously.

For all predictor variables, we introduced a lag of seven days to account for an incubation period [163] and time to reporting. First, we considered a negative binomial model with all of the 20 aforementioned risk factors and the two corresponding infection counts from the previous two weeks of the district as predictors. Adapting the model specification of Equation 2.3, our model was then defined as:

$$\begin{aligned} &I_t|\mathcal{F}_{t-1} \sim NegBin(\lambda_t, \psi) \\ &\log(\lambda_t) = \beta_0 + \beta_1 \log(I_{t-1}+1) + \beta_2 \log(I_{t-2}+1) + \eta_1 X_{t-1,1} + \ldots + \eta_{20} X_{t-1,20}, \end{aligned}$$

where I_t is the number of infected in week t, $\mathcal{F}_{t-1}$ is the history of the process until

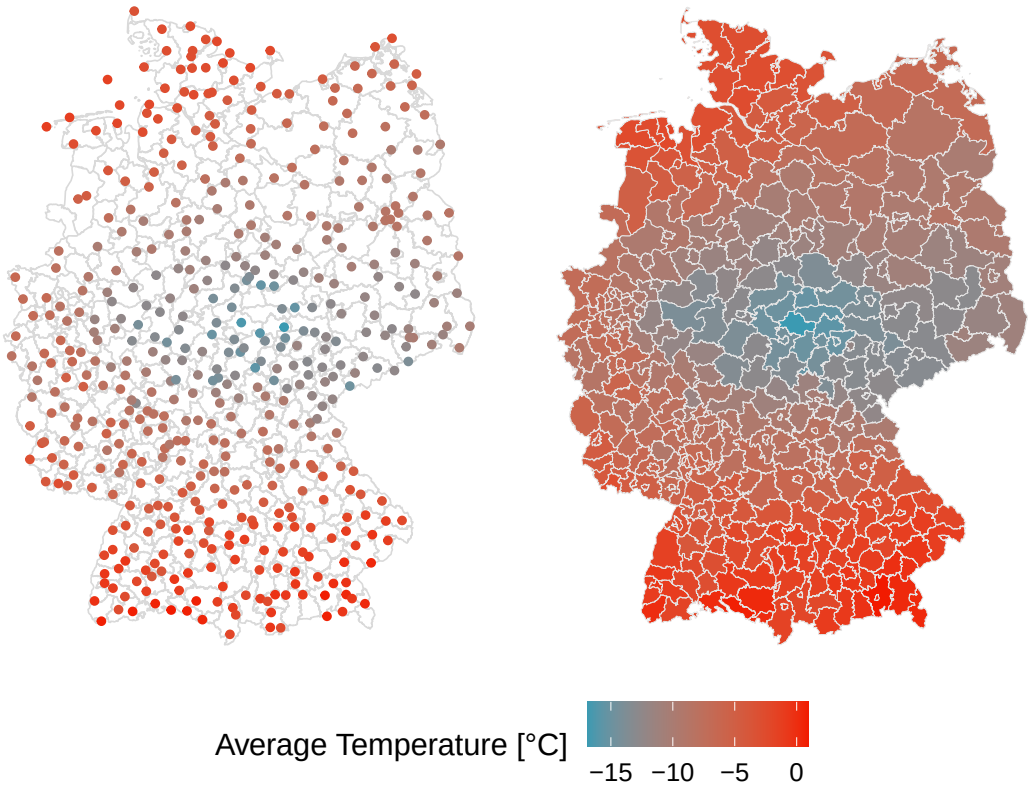

Figure 3.3: Example interpolation of average temperatures in the districts based on weather station measurements on February 9, 2021. On the left, weather stations are indicated as coloured points, showing the average temperature measured at the respective station. On the right, the interpolated values for the respective districts are displayed.

$t-1$, and $X_{t-1,i}$ is the value of the ith covariate in week $t-1$ with $i = 1, \ldots, 20$ and $t = 3, \ldots, 106$. For each of the four analysed districts we created a separate model. As mentioned, we used the R package `tscount` to fit the model.

In order to select the most relevant predictors for the models and avoid overfitting, we implemented a step-wise selection algorithm. We used the QIC as selection criterion and considered in each step the models, where one of the currently included covariates were excluded or one of the previously excluded covariates were included again. The autoregressive factors were not considered for exclusion. The model with the smallest QIC was then selected for the next step. Each of the potential new models was fitted as a negative binomial model, however, if the overdispersion parameter was estimated at zero such that the variance equalled the mean, we refitted the model as a Poisson model. The algorithm stopped when all of the considered models had a higher QIC than the current model.

To also determine the predictive power of our method, we considered an adapted approach. We defined the data of the first year as training data, such that the data of the 51 weeks between February 24, 2020, and February 21, 2021, were used for variable selection and model fitting as described above. Then, for each of the following weeks, we calculated one-week-ahead predictions and 95% prediction intervals based on this selected model, which was refitted on all previous weeks each time.

3.4.3 Results

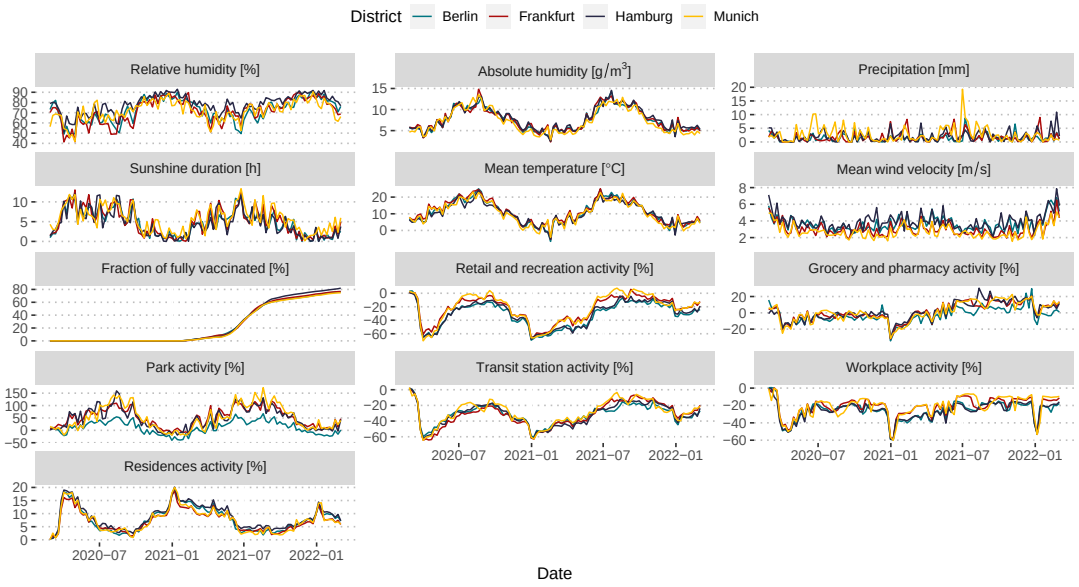

Figure 3.4: Weekly values of predictor variables between February 24, 2020, and March 6, 2022, for the districts Berlin, Frankfurt, Hamburg, and Munich.

Looking at the risk factors over the analysis period, some weather variables had very fluctuating movements of their weekly values without strong seasonal effects, such as relative humidity, precipitation, wind velocity (Figure 3.4). On the other hand, temperature, absolute humidity, sunshine duration, as well as park activity were very season-dependent. The vaccination rate increased strongly in autumn of 2021 and kept steadily increasing after that, however, without major differences between the four analysed cities. The disruptive effect on activity patterns due to the first lockdown between March and May, 2020, the second lockdown between December 2020 and January 2021, and the holiday season at the end of each year is apparent in the changes regarding retail, recreation, transit stations, workplaces, and residences.

For the association analysis regarding the full time period, the stepwise selection

	Incidence rate ratio (95% confidence interval)			
	Berlin	Frankfurt	Hamburg	Munich
Lockdown 1	0.28 (0.27, 0.29)	0.37 (0.19, 0.71)		0.65 (0.41, 1.05)
Lockdown 2	0.72 (0.71, 0.73)		0.65 (0.64, 0.67)	0.78 (0.51, 1.18)
Lockdown soft			0.76 (0.74, 0.77)	
Lag 2 infections	0.99 (0.98, 1.01)	0.75 (0.49, 1.14)	0.86 (0.85, 0.87)	0.72 (0.55, 0.94)
Absolute humidity	1.04 (1.02, 1.05)		0.91 (0.90, 0.92)	0.82 (0.63, 1.05)
Transit station activity	1.03 (1.02, 1.03)	0.94 (0.90, 0.99)	0.95 (0.95, 0.95)	
Grocery and pharmacy activity		0.99 (0.96, 1.01)	0.99 (0.99, 0.99)	0.95 (0.92, 0.97)
Precipitation	0.97 (0.97, 0.98)		0.98 (0.98, 0.98)	
Park activity	0.99 (0.99, 0.99)		1.00 (1.00, 1.00)	
Workplace activity			1.01 (1.01, 1.01)	
Temperature	1.01 (1.00, 1.01)		1.02 (1.01, 1.03)	1.03 (0.92, 1.16)
Relative humidity	1.01 (1.01, 1.01)	1.01 (0.99, 1.03)	1.04 (1.04, 1.04)	1.02 (0.99, 1.06)
Retail and recreation activity	1.02 (1.02, 1.02)	1.02 (1.00, 1.04)	1.04 (1.04, 1.04)	1.03 (1.02, 1.05)
Sunshine duration	1.05 (1.04, 1.05)		0.99 (0.98, 0.99)	1.05 (0.97, 1.14)
Autumn	1.08 (1.06, 1.10)	1.31 (0.92, 1.86)	0.76 (0.75, 0.78)	
Residences activity	1.20 (1.19, 1.21)	0.91 (0.78, 1.06)		
Winter	1.10 (1.08, 1.13)	1.13 (0.71, 1.79)	0.80 (0.79, 0.82)	1.21 (0.85, 1.73)
Neighbourhood infections	0.77 (0.76, 0.78)			1.64 (0.95, 2.86)
Spring	1.29 (1.26, 1.32)	1.49 (0.88, 2.52)		1.36 (0.78, 2.40)
Lag 1 infections	2.72 (2.69, 2.75)	2.72 (1.74, 4.25)	2.72 (2.68, 2.76)	1.83 (1.02, 3.30)
Vaccination rate	2.22 (2.17, 2.28)	5.89 (1.92, 18.06)		2.30 (1.22, 4.36)

Table 3.1: Exponentiated coefficient estimates, called incidence rate ratios, with 95% confidence intervals for infection count models of Berlin, Frankfurt, Hamburg, and Munich. Poisson models were used for Berlin and Hamburg, negative binomial models for Frankfurt and Munich. Empty entries indicate that the corresponding variable was not included in the model of the respective district. Wind speed was not selected in any model.

methods selected Poisson models for the districts Berlin and Hamburg. Negative binomial models were selected for Frankfurt and Munich. In general, the Poisson models included more covariates than the negative binomial models, incorporating almost all of the potential predictors (Table 3.1). Furthermore, almost all confidence intervals of the covariate estimates indicated significance in the Poisson case, while barely any significant ones were identified in the negative binomial case.

Vaccination rate, infection count of the previous week, retail and recreational activity, as well as relative humidity had a consistently positive association with infections of a district when included in the model. On the other hand, negative associations emerged for lockdowns, infection counts from two weeks ago, activity in grocery stores and

pharmacies, and precipitation. Large differences between the incidence rate ratios (IRR) appeared for infections in neighbouring districts between Berlin (IRR: 0.77, 95% confidence interval (CI): 0.76-0.78) and Munich (IRR: 1.64, CI: 0.95-2.86). Similar differences were also visible for other covariates, including absolute humidity, transit station activity, sunshine duration, and seasonal indicators.

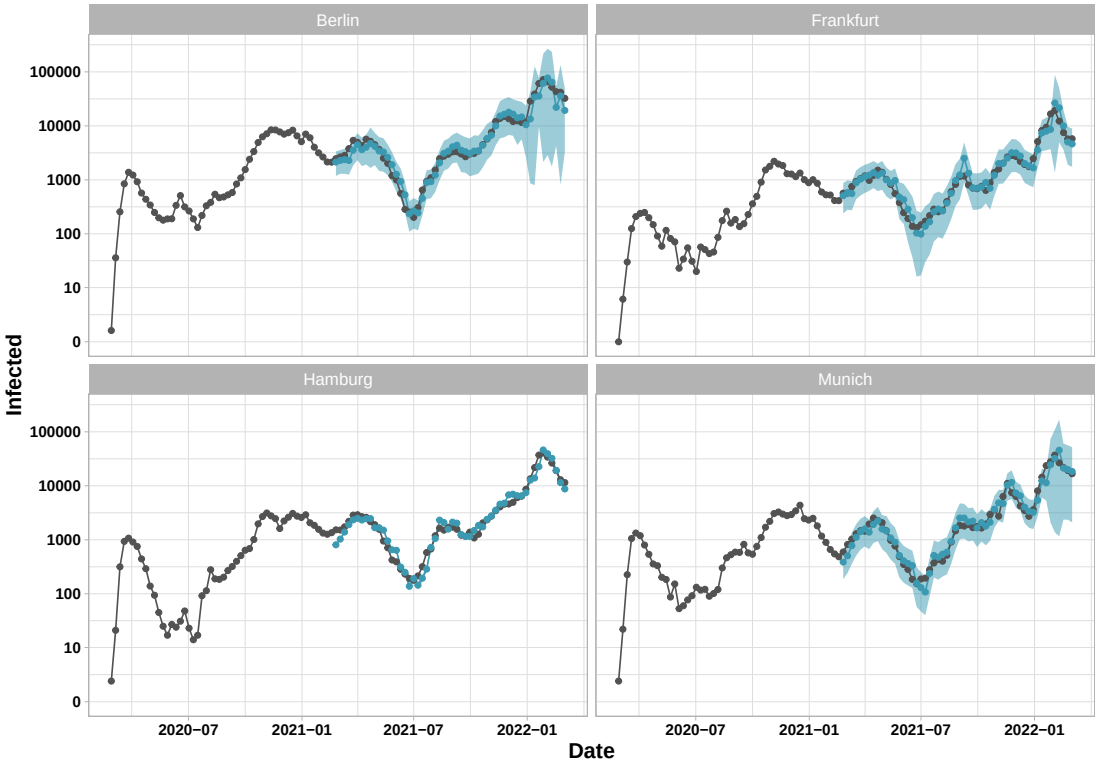

Figure 3.5: Infection curves (black) and one-week-ahead predictions (blue) with 95% prediction intervals for the districts Berlin, Frankfurt, Hamburg, and Munich.

For the forecast models, where the selection method was applied to the data of the first 51 weeks, a Poisson model was selected only for Hamburg. Figure 3.5 shows one-week-ahead predictions and corresponding 95% prediction intervals, which illustrate the difference in the predictions between Poisson and negative binomial models. The SMDAPE was smallest for the Poisson model in Hamburg with 15.0% compared to 17.0%, 15.6%, and 15.7% in Berlin, Frankfurt, and Munich, respectively. Thus, the forecasts were similarly accurate for all districts. However, only 20% of the prediction intervals covered the observed values for Hamburg, while the negative binomial prediction intervals contained 53 of the predicted 54 weekly infection counts (98%) for each of the other districts.

3.4.4 Limitations

Multiple limitations and options for future improvement became apparent in this study. Selecting an appropriate distribution for modelling infection counts was one of the most crucial points of the analysis. Poisson models have the property that their variance is equal to their mean parameter. Thus, Poisson models may include more predictor variables and their coefficient estimates may have a higher significance compared to negative binomial models to compensate for this restriction. Thus, the possibility to compare results of Poisson and negative binomial models can be very limited and coefficient significance has to be interpreted with caution. To emphasize this conclusion, we adapted our inferential approach by substituting the selected covariates by random samples from normal distributions with no associations to the outcomes. The Poisson models indicated significant associations to almost all of the new covariates, while the negative binomial models barely found any association (Appendix C.1). The fact that many publications relied on the Poisson distribution and derived seemingly impactful associations [164–166] highlights the importance to select appropriate models with due caution. Our Poisson model used for forecasting had very small prediction intervals and, although their mean estimates were close to the actual values, confidence intervals rarely covered the observed infection counts, further emphasizing the critical concern with any inappropriate use of such models.

Despite a general estimate of incubation time, our definition of a seven-day lag for all covariates was arbitrary, at least to some degree. We only had data available for the incubation period, not the time until cases were reported, and incubation periods varied greatly between COVID-19 variants [167]. Future work may include the data-driven selection of appropriate time lags or a distributed lag model as well as virus-variant-specific and covariate-specific adaption of the lag. Another level of added precision that we did not include was the age and gender information. The differences in infection counts with respect to these two variables explored in Section 1.2 indicated some relevance. These implementations, however, would increase the risk of overfitting and may superficially lead to significance when estimating coefficients. Applying these changes to the model would therefore require careful and thorough implementation.

Finally, the analysis could be further improved if data with higher resolution and accuracy were available. This could also improve the forecasting performance, which had acceptable point estimates for one-week-ahead predictions, which would, however, lead to large deviations for multiple-week-ahead predictions and already had quite large prediction intervals at various time points, at least for the negative binomial models. For example, mobility data was only available on a state level. Non-pharmaceutical

interventions were far more numerous and locally restricted than the three nationwide lockdowns we included in our analysis would suggest. Unfortunately, data in that regard with high enough spatial resolution and, in particular, enough reliability were not available (see also Appendix C.2). Since the analysis was based on aggregated data on a district level, any inference from the results had to be considered with caution and in some instances could not substitute a patient-level analysis. An example for this is the incidence rate ratio of the vaccination rate, which indicated that a higher vaccination rate would lead to more COVID-19 infections. However, a multitude of patient-level analyses showed that vaccinations decreased the infection risk [168–170], thus indicating that high infection numbers only coincided with higher vaccination rates, in particular towards the end of 2021 and beginning of 2022 (Figures 3.4 and 3.5), and were not caused by them. More infections, but less severity, can be explained in part by increased prevalence of different virus variants [171, 172], which we could not consider due to lacking data availability. In conclusion, future analyses could be improved substantially by incorporating time-dependent virus variant predominance and other essential information.

3.4.5 Conclusion

We performed a longitudinal analysis of weekly COVID-19 infection counts in four German districts considering a time frame of two years after the pandemic started. In contrast to many studies, we incorporated autoregressive features and neighbouring district infections and allowed for Poisson as well as negative binomial models, while accounting for potentially relevant covariates. Coefficient estimates showed large differences between Poisson and negative binomial models. Poisson models had the tendency to overestimate coefficient significance considerably.

Nonetheless, the negative associations of lockdowns and activity in grocery stores and pharmacies with infection counts became apparent. Similarly, positive associations were found for retail and recreational activities and the number of infections in the previous week. Finally, we showed the predictive capabilities of our approach, which produced accurate one-week-ahead predictions, however, prediction intervals in the case of the Poisson models were much too small. Prediction intervals of the negative binomial models were large but accurate.

Here we preferred frequentist methods for our longitudinal analysis, yet Bayesian methods should not be discarded and have already demonstrated the ability to model infectious diseases [173, 174]. Another consideration that could be incorporated in the future is dynamic autoregression. Since behaviour and spread of a disease can vary, a

model could adapt accordingly by relaxing the linear model to a piecewise constant linear model or even a separate model for each time point such that the model parameters can be fully time-dependent [175].

Our frequentist approach enabled us to consider a large number of covariates and a data-based approach to select a relevant subset in a computationally manageable way. However, we also saw that the wide range of analysis techniques carries the risk of easy misuse and misleading results. Therefore, a critical evaluation of the applied methods and the resulting outcomes is essential, especially in the analysis of time series, due to the many possibilities and ease of adjusting potential covariates – intentionally or unintentionally – to the outcome.

Chapter 4

Discussion

We have already discussed our individual results in detail, from which we were able to derive very specific goals for the near future. We are currently working on further risk tools, extending from kidney to pancreas and liver, based on thousands of transplants in the Eurotransplant region. We also found that the relevance and appropriateness of approaches to account for clustering is unclear for large sample sizes but small groups. We plan to derive clear recommendations by simulating comparable circumstances in survival analyses. However, we would also like to point to possible further development in a larger context and mention some other important findings from our analyses.

There is a plethora of possibilities that enables the modern statistician to analyse temporal information in all its facets. The broad spectrum of disciplines, such as econometrics, signal processing, political sciences, and biometry, having a vested interest in findings from this kind of data, lead to very different expectations and therefore available methods. Curve fitting, decomposition, prediction, classification, and segmentation are just a few of the general topics that cover the analysis of time series. All of them utilize many methods, including harmonic models, Box-Jenkins forecasting, wavelets, and deep neural networks, to understand the temporal dimension in all its complexity [176, 177]. Thus, there are many ways to start and expand analyses of temporal data.

But also the progression to a higher focus on spatial information is often an option. Spatial analyses, particularly native to areas such as astronomy, geography, and biology, have their own questions and methods, including correlation analysis, interpolation, and spatial regression. However, spatial information is often not as structured as other data due to irregular measurement systems, non-uniform administrative boundaries, multi-dimensionality, alternative projections, and inhomogeneities of earth and universe. As a result, large fields of application, such as geographic information systems, evolved in order to store, manage, visualize, as well as analyse geospatial data [178].

Therefore, it is not surprising that space and time are often explored independently for simplicity and clarity. Yet, focused analyses based on the connection of space and time

might be inevitable to elevate the discovery process. Compared to separate analyses, the development of statistical methods for the combined analysis of space and time has not made as much progress for a long period of time despite increasing abundance of spatio-temporal data. But, there have been a number of investigations of spatio-temporal processes particularly in fields such as spatial epidemiology, for example, reviewed by Elliott & Wartenberg [179]. Furthermore, comprehensive statistical methodology for spatio-temporal processes, as described by Cressie & Wikle [180] among others, has recently seen a surge in application [181–188]. Due to the strong spatio-temporal nature of the spread and major impact of COVID-19, this trend has further intensified.

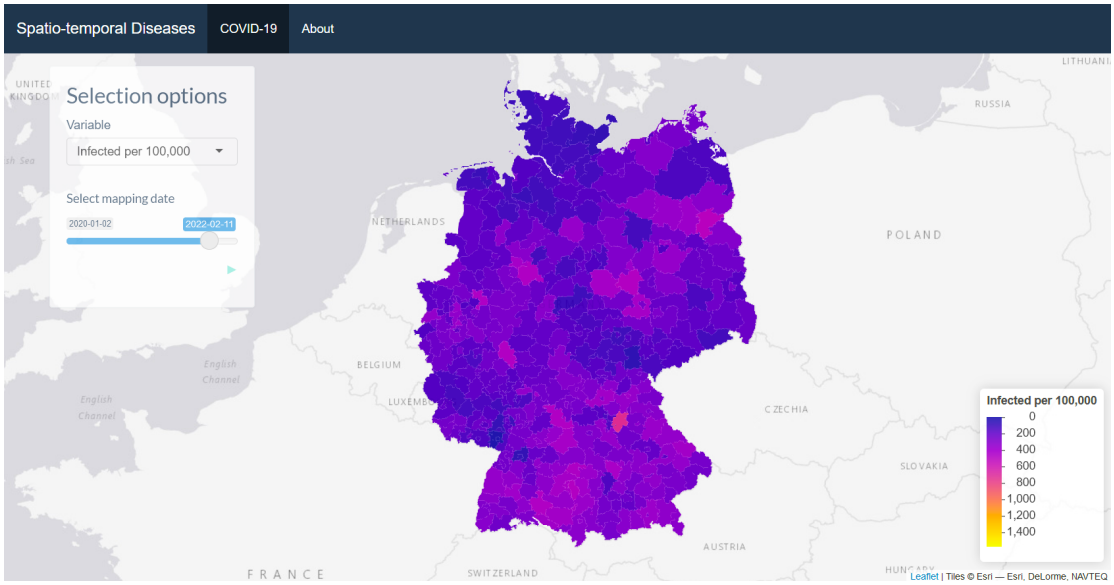

Figure 4.1: Screenshot of our interactive COVID-19 metrics tool displaying infection counts per 100,000 inhabitants for German districts on February 11, 2022.

Even though our analyses already included limited efforts in linking these two dimensions, there is much more potential to intensify and implement these efforts in the future. However, with higher complexity, there is a risk that results will be misleading due to inappropriate methods. Our analyses have shown that seemingly minuscule changes in model specification can dramatically change the message of the results in such a context. For example, our findings have demonstrated that the selection of covariates is one of the crucial steps in such analyses, which is often non-trivial and much debated for good reason [189–191]. Therefore, methods must be selected and applied with great care, and results should be questioned and verified. Especially for longitudinal count data, we showed that the selection of a suitable model and extensive validation are vital to avoid misleading results and to extract reliable conclusions.

Considering our analyses on kidney transplants, we believe that a stronger focus on regions and even more local approaches are necessary, as even at centre-specific levels there are often very different processes and conditions present. These efforts will be severely limited due to privacy regulations, nonetheless, our approach of focusing model validation on cross-country comparisons can be considered a very small step in this direction. As another example, current public sources on COVID-19, which are abundantly available, often consider spatial and temporal information only separately [192–195]. The principle of linking space and time more closely should therefore be extended not only to modelling approaches but also to general information efforts. Thus, we have developed an intuitive tool to display various metrics such as the number of infections or cumulative deaths for all German districts on a daily scale, for specific dates, but also as a sequence since the beginning of the pandemic (Figure 4.1). Although we have highlighted a few pitfalls on the path to reliable results, we hope that this is just the beginning of further efforts to consider time and space jointly to make future findings not only faster and easier but also more substantial and accurate.

Appendix A

List of Acronyms

AIC	Akaike Information Criterion	52
AUC	area under the receiver operating characteristic curve	20
BIC	Bayesian Information Criterion	19
ET	Eurotransplant	56
GEE	generalized estimating equations	52
HMC	Hamiltonian Monte Carlo	42
IHME	Institute for Health Metrics and Evaluation	29
INGARCH	integer-valued generalized autoregressive conditional heteroscedasticity	38
JAGS	Just Another Gibbs Sampler	42
MCMC	Markov Chain Monte Carlo	38
MICE	multiple imputation by chained equations	19
NPI	non-pharmaceutical interventions	29
QIC	quasi-likelihood under the independence model criterion	52
RKI	Robert Koch-Institute	7
SEIR	susceptible-exposed-infectious-recovered	24
SMDAPE	Symmetric Median Absolute Percentage Error	53
UT	University of Texas	32
WHO	World Health Organization	5

Appendix B

Remarks on Implementation

B.1 `tsglm` output

The following shows R output of the `tsglm` function used in section 2.3.1 for a sample of the COVID-19 data of Munich.

```
Call:
tsglm(ts = y,
model = list(past_obs = 1),
link = "log",
distr = "poisson")

Coefficients:
             Estimate  Std.Error  CI(lower)  CI(upper)
(Intercept)     0.674     0.0461      0.583      0.764
beta_1          0.842     0.0104      0.821      0.862
Standard errors and confidence intervals (level =  95 %)
obtained by normal approximation.

Link function: log
Distribution family: poisson
Number of coefficients: 2
Log-likelihood: -2010.577
AIC: 4025.153
BIC: 4032.047
QIC: 4025.153
```

B.2 Metropolis sampler code

The following shows the Metropolis R code used in section 2.3.1 for the Poisson model of the COVID-19 data of Munich.

```
###################################
# Factor of the log probability
log_prob_f <- function(coefs_old, coefs_new, sumi, sumli, ↷
   ↪ exp_lagged, phi){
  facto <- (coefs_old[1]^2 - coefs_new[1]^2)/(2*phi[1]) + ↷
     ↪ (coefs_old[2]^2 - coefs_new[2]^2)/(2*phi[2]) +
  sumi*(coefs_new[1] - coefs_old[1]) + sumli*(coefs_new[2] - ↷
     ↪ coefs_old[2]) +
  exp(coefs_old[1])*sum(exp_lagged^coefs_old[2]) - ↷
     ↪ exp(coefs_new[1])*sum(exp_lagged^coefs_new[2])
  return(facto)
}

###################################
# The proposal function
Proposal_opt <- function(coefs, Sigma){
  return(mvrnorm(n = 1, mu = coefs, Sigma = Sigma))
}

###################################
# Sampler
MetropolisSampler_opt <- function(init, iterations, ↷
   ↪ predictors, output, phi, Sigma){
  # Initialize array to save chain values, with acceptance info
  if(is.function(init)){startvalues <- ↷
     ↪ init()}else{startvalues <- init}
  chain <- array(dim=c(iterations, k + 2))
  chain[1,] <- c(1, TRUE, startvalues)
  current_coefs <- startvalues

  # Initial calculations
  sumi <- sum(output)
  sumli <- sum(output*predictors[,2])
  exp_lagged <- exp(predictors[,2]) # take exp of log(I_{t-1} ↷
     ↪ + 1)

  # Run loop
```

```
  for (i in 2:iterations){
    # Generate theta from proposal
    suggested <- Proposal_opt(coefs = current_coefs, Sigma)
    # Compute acceptance probability
    log_prob <- min(0, log_prob_f(coefs_old=current_coefs,
    coefs_new=suggested,
    sumi=sumi, sumli=sumli,
    exp_lagged=exp_lagged, phi=phi))
    # Decision to accept Y
    if(log(runif(1))<log_prob){
      chain[i,] <- c(i, TRUE, suggested)
      # Change values of current step
      current_coefs <- suggested
    }else{
      chain[i,] <- c(i, FALSE, current_coefs)
    }
  }
  return(as.data.frame(chain))
}
```

B.3 Stan code

The following shows the Stan script used in section 2.3.1 for the Poisson model of the COVID-19 data of Munich.

```
data
{
  // Number of observations
  int<lower=0> n;
  //Number of beta parameters
  int<lower=0> k;
  // Covariates
  matrix[n, k] X;
  // Outcome
  int<lower=0> y[n];
  // Given variances
  vector[k] phi;
}

parameters
```

```
{
  vector[k] beta;
  // matrix[k, k] Sigma;
}

transformed parameters
{
  vector[n] lambda;
  lambda = exp(X * beta);
}

model
{
  // Priors - no need to specify non-informative priors
  for (i in 1:k)
  {
    beta[i] ~ normal(0, phi[i]);
  }
  // Likelihood part
  y ~ poisson(lambda);
}
```

With the corresponding variables set, this script can be employed in R by running the below code.

```
# Library and Stan properties
library(rstan)
options(mc.cores = parallel::detectCores())
rstan_options(auto_write = TRUE)

# Data inputs
data <- list(y=y, X=X, k=k, n=n, phi=phi)

# Run sampling
stan_sim <- stan(file = filepath,
                 data = data,
                 pars = "beta",
                 iter = iterations,
                 warmup = burnin,
                 chains = n_chains,
                 init = starting_values)
```

B.4 Simulations

The figures in this section visualize auxiliary results with respect to sample values and coefficient estimates of Poisson simulations described in chapter 2.3.1.

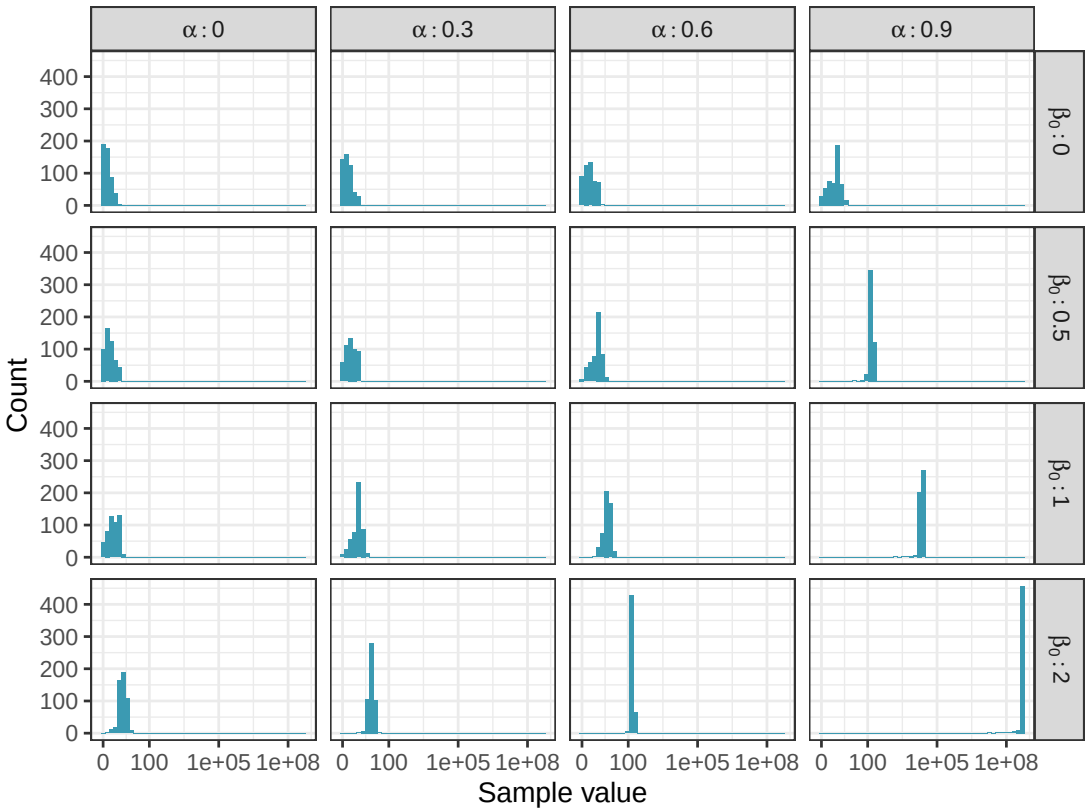

Figure B.1: Histogram of single simulations comprising 500 observations distributed by the autocorrelated Poisson model defined in Equation 2.2 with different combinations of β_0 and α.

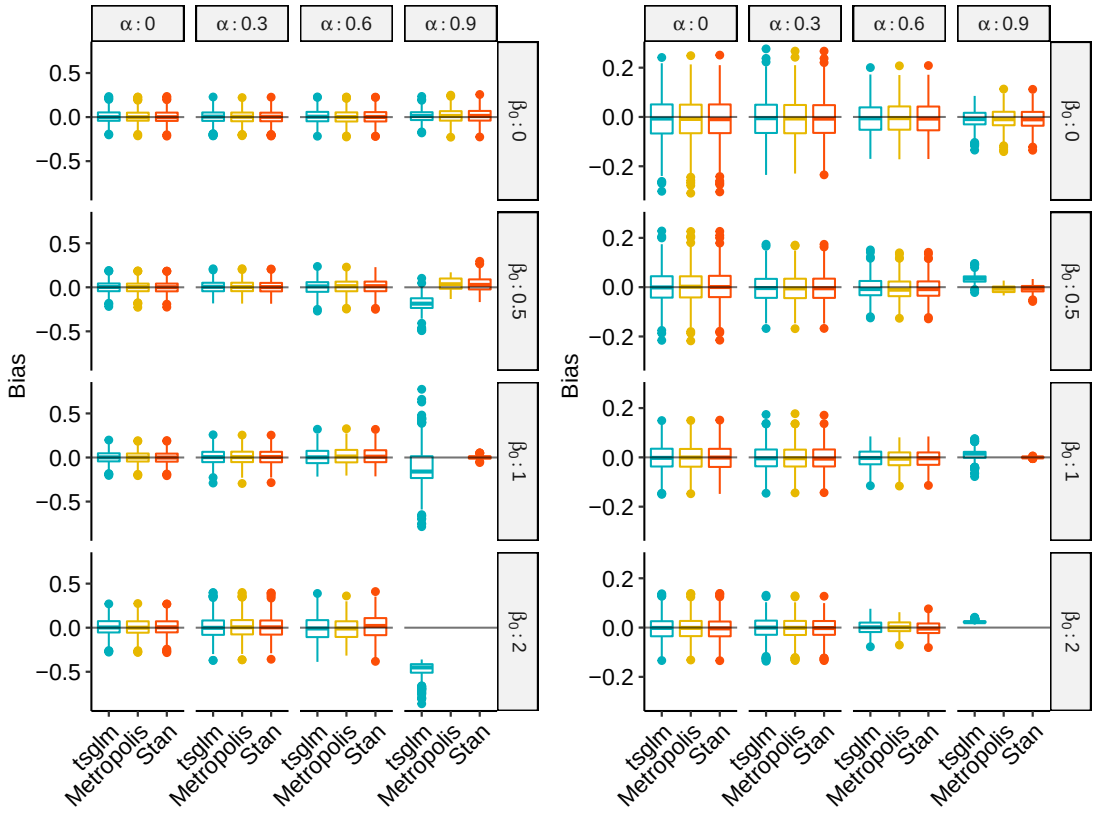

Figure B.2: Coefficient estimate bias of β_0 (left) and α (right) in simulations comprising 500 observations distributed by the autocorrelated Poisson model defined in Equation 2.2 with different combinations of β_0 and α values. 100 simulations were performed for each combination.

Appendix C

Remarks on application

C.1 Random covariates in Poisson models

	Incidence rate ratio (95% confidence interval)			
	Berlin	Frankfurt	Hamburg	Munich
Random5	0.98 (0.98, 0.98)	1.03 (1.00, 1.07)	1.00 (1.00, 1.00)	1.00 (0.94, 1.06)
Random4	1.00 (1.00, 1.00)	1.01 (0.96, 1.05)	0.97 (0.97, 0.97)	0.98 (0.90, 1.06)
Random3	0.98 (0.98, 0.98)	0.98 (0.91, 1.05)	0.97 (0.97, 0.97)	1.02 (0.92, 1.13)
Random2	0.98 (0.98, 0.98)	0.98 (0.91, 1.05)	0.95 (0.95, 0.95)	0.97 (0.85, 1.11)
Random1	1.02 (1.02, 1.02)	0.91 (0.78, 1.08)	1.05 (1.04, 1.05)	1.01 (0.72, 1.44)
Lag 2 infections	0.96 (0.95, 0.97)	0.94 (0.54, 1.66)	0.90 (0.89, 0.91)	0.95 (0.45, 2.01)
Lag 1 infections	2.71 (2.69, 2.74)	2.72 (1.50, 4.93)	2.71 (2.68, 2.74)	2.63 (1.24, 5.57)

Table C.1: Exponentiated coefficient estimates, which are incidence rate ratios, with 95% confidence intervals for infection count models for Berlin, Frankfurt, Hamburg, and Munich. Poisson models were used for Berlin and Hamburg, negative binomial models for Frankfurt and Munich. Random1 through Random5 are samples from a normal distribution with mean 0 and standard deviation 1, 2, 3, 4, and 5, respectively.

The variance of Poisson distributions equals the mean. This means that Poisson models estimate a significant effect of covariates more easily to compensate for observations with higher variance than, for example, negative binomial models. To exemplify this, we adapted the models from Chapter 3.4 modelling the weekly COVID-19 infections in four German districts from February 24, 2020, until March 6, 2022. There, we used as predictors the two previous weekly counts and a data-based selection of covariates. We sampled from five normal distributions with mean zero and standard deviation 1, 2, 3, 4, and 5, respectively, and substituted the selected covariates with them. The coefficient estimates and their confidence intervals of the newly fit models are shown in C.1, where we used Poisson models for the two districts Berlin and Hamburg and negative binomial

models for Frankfurt and Munich. While the negative binomial models did not indicate a significant association for almost any of the randomly sampled covariates, the Poisson models did so for almost all of them.

C.2 Intervention data

In order to integrate the effect of interventions into our models we sought a public database collecting the type and time frame of these measures on a regional level. Even though this kind of information is fundamental for reproducibility and comparability in analyses exploring intervention effects, the supply of such databases is very limited. One of the few available is the CoronaNet Research Project database [196]. The CoronaNet Research Project compiles government responses based on the input of more than 450 research assistants from around the world. Unfortunately, the data is by far not complete or fully up-to-date, thereby implying false intervention effects. A selection of extracted, processed, and aggregated data on a few of the measures is shown in Figure C.1.

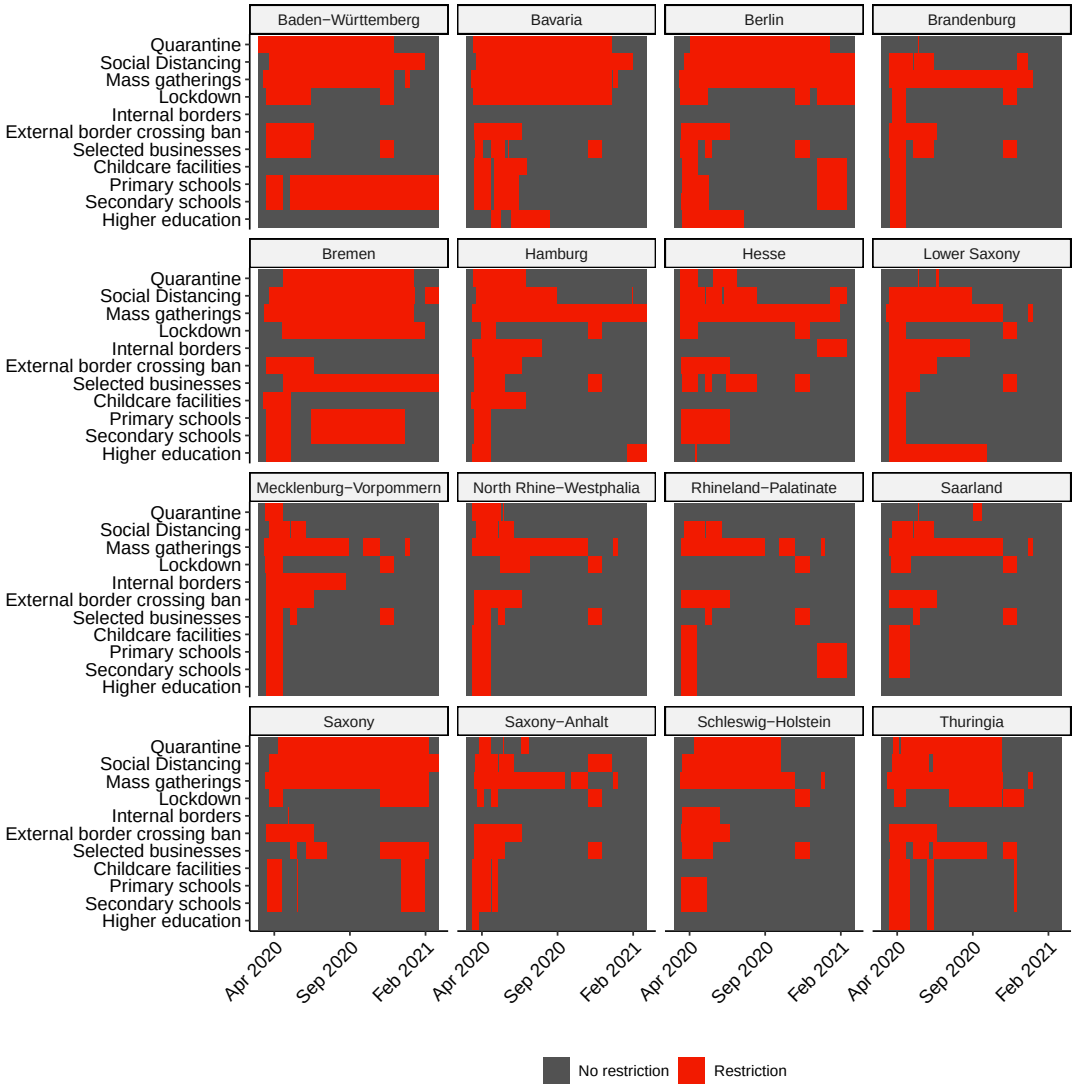

Figure C.1: Processed intervention data extracted on 2021-03-24 from the CoronaNet Research Project database [196] between 2020-03-01 and 2021-02-28.

List of Figures

1.1 Cumulative infections per 100,000 inhabitants in German districts for four different dates. The top left figure shows the German state names, while the others highlight notable districts. 6

1.2 Daily infections and deaths in Germany from January 28 to August 31, 2020. 8

1.3 Daily reported number of infections (left) and deaths (right) from week of March 3 (light) to May 30 (dark). 9

1.4 Boxplots of reported infections and deaths across the German districts during the period January 28 to August 31, 2020; less than 1% of the respective counts with unknown gender or age are not shown. 10

2.1 Diagram of time intervals in infectious diseases (based on Lipsitch *et al.* [95] and Porta [96]). 26

2.2 Major routes of outbound air and train travel originating from Wuhan during the Spring Festival travel season, 2019. Darker and thicker edges represent greater numbers of passengers (from Wu *et al.* [86]). 28

2.3 Weekly excess death rates during the 1918 influenza pandemic from September 8, 1918, through February 22, 1919 (from Markel *et al.* [107]). 30

2.4 Visitation patterns in Texas visualized as relative changes to pre-pandemic baselines for various points of interest (from Woody *et al.* [84]). The vertical line showing when the state death rate reached 3 per 10 million residents. 33

2.5 Introduction dates of interventions (from Flaxman *et al.* [117]). 36

2.6 Time series of daily COVID-19 infections in Munich from 2020-02-24 to 2021-03-16 (black) in comparison to Poisson models (coloured) with different parameter combinations each with 10 simulations. 38

2.7 Trace plots of Metropolis sampler and Stan fitting Poisson model of COVID-19 infections in Munich. Only the first 500 draws are shown, which are discarded as burn-in. 47

2.8 Coverage rates in percent of the confidence interval in the case of `tsglm` and credible intervals in the case of Metropolis and Stan for simulated data. One simulation comprises 500 observations distributed by the autocorrelated Poisson model defined in Equation 2.2 with different combinations of β_0 and α values. 500 simulations were performed for each combination. Gray cells indicate non-convergence for all simulations. . . 48

3.1 Screenshot of the online risk calculator for kidney transplants KTOP available at `riskcalc.org/ktop`. 59
3.2 The four German districts for which longitudinal analyses were performed for the time period between February 24, 2020, and March 6, 2022, and their neighbouring districts. 61
3.3 Example interpolation of average temperatures in the districts based on weather station measurements on February 9, 2021. On the left, weather stations are indicated as coloured points, showing the average temperature measured at the respective station. On the right, the interpolated values for the respective districts are displayed. 64
3.4 Weekly values of predictor variables between February 24, 2020, and March 6, 2022, for the districts Berlin, Frankfurt, Hamburg, and Munich. 65
3.5 Infection curves (black) and one-week-ahead predictions (blue) with 95% prediction intervals for the districts Berlin, Frankfurt, Hamburg, and Munich. 67

4.1 Screenshot of our interactive COVID-19 metrics tool displaying infection counts per 100,000 inhabitants for German districts on February 11, 2022. 72

B.1 Histogram of single simulations comprising 500 observations distributed by the autocorrelated Poisson model defined in Equation 2.2 with different combinations of β_0 and α. 81
B.2 Coefficient estimate bias of β_0 (left) and α (right) in simulations comprising 500 observations distributed by the autocorrelated Poisson model defined in Equation 2.2 with different combinations of β_0 and α values. 100 simulations were performed for each combination. 82

C.1 Processed intervention data extracted on 2021-03-24 from the CoronaNet Research Project database [196] between 2020-03-01 and 2021-02-28. 85

List of Tables

1.1 COVID-19 metrics by age and gender in Germany until August 31, 2020. Cases with unknown age or gender are not shown. 11

2.1 Selection of infectious disease models. NegBin corresponds to the negative binomial distribution, GAMM to generalized additive mixed model, SEIR to susceptible-exposed-infectious-recovered model. 25

2.2 Results for β_0 and α across Bayesian sampling methods with respect to daily COVID-19 infection numbers in Munich from 2020-02-24 to 2021-03-16. SE is the standard error, $\widehat{R}$ is an indicator for convergence with values < 1.01 indicating convergence. For the Metropolis sampler the function `Rhat` from the package `rstan` was used to calcuate $\widehat{R}$, the other methods reported it as standard output of their function. In the case of WinBUGS the runtime only comprises the access time from R to the program, since it requires user input. 45

3.1 Exponentiated coefficient estimates, called incidence rate ratios, with 95% confidence intervals for infection count models of Berlin, Frankfurt, Hamburg, and Munich. Poisson models were used for Berlin and Hamburg, negative binomial models for Frankfurt and Munich. Empty entries indicate that the corresponding variable was not included in the model of the respective district. Wind speed was not selected in any model. . . . 66

C.1 Exponentiated coefficient estimates, which are incidence rate ratios, with 95% confidence intervals for infection count models for Berlin, Frankfurt, Hamburg, and Munich. Poisson models were used for Berlin and Hamburg, negative binomial models for Frankfurt and Munich. Random1 through Random5 are samples from a normal distribution with mean 0 and standard deviation 1, 2, 3, 4, and 5, respectively. 83

Bibliography

1. Zeger, S. L., Irizarry, R. & Peng, R. D. On time series analysis of public health and biomedical data. *Annual Review of Public Health* **27,** 57–79 (2006).
2. Chadban, S. J. *et al.* Prevalence of kidney damage in Australian adults: The AusDiab kidney study. *Journal of the American Society of Nephrology* **14,** S131–S138 (2003).
3. Coresh, J., Astor, B. C., Greene, T., Eknoyan, G. & Levey, A. S. Prevalence of chronic kidney disease and decreased kidney function in the adult US population: Third National Health and Nutrition Examination Survey. *American Journal of Kidney Diseases* **41,** 1–12 (2003).
4. Bikbov, B. *et al.* Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **395,** 709–733 (2020).
5. Liyanage, T. *et al.* Worldwide access to treatment for end-stage kidney disease: a systematic review. *The Lancet* **385,** 1975–1982 (2015).
6. Stel, V. S. *et al.* Time trend in access to the waiting list and renal transplantation: a comparison of four European countries. *Nephrology Dialysis Transplantation* **27,** 3621–3631 (2012).
7. Meier-Kriesche, H. & Kaplan, B. Waiting time on dialysis as the strongest modifiable risk factor for renal transplant outcomes: A Paired Donor Kidney Analysis. *Transplantation* **74,** 1377–1381 (2002).
8. Kasiske, B. L., Snyder, J. J., Gilbertson, D. & Matas, A. J. Diabetes mellitus after kidney transplantation in the United States. *American Journal of Transplantation* **3,** 178–185 (2003).
9. Engels, E. A. *et al.* Spectrum of cancer risk among US solid organ transplant recipients. *Journal of the American Medical Association* **306,** 1891–1901 (2011).
10. Vajdic, C. M. *et al.* Cancer incidence before and after kidney transplantation. *Journal of the American Medical Association* **296,** 2823–2831 (2006).

11. Port, F. K., Wolfe, R. A., Mauger, E. A., Berling, D. P. & Jiang, K. Comparison of survival probabilities for dialysis patients vs cadaveric renal transplant recipients. *Journal of the American Medical Association* **270,** 1339–1343 (1993).

12. Wolfe, R. A. *et al.* Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *New England Journal of Medicine* **341,** 1725–1730 (1999).

13. Tonelli, M. *et al.* Systematic review: kidney transplantation compared with dialysis in clinically relevant outcomes. *American Journal of Transplantation* **11,** 2093–2109 (2011).

14. Nicoletto, B. B. *et al.* Effects of obesity on kidney transplantation outcomes: a systematic review and meta-analysis. *Transplantation* **98,** 167–176 (2014).

15. Woo, Y. M. *et al.* Early graft function and patient survival following cadaveric renal transplantation. *Kidney International* **55,** 692–699 (1999).

16. Traynor, C. *et al.* Twenty-year survivors of kidney transplantation. *American Journal of Transplantation* **12,** 3289–3295 (2012).

17. McAdams-DeMarco, M. *et al.* Frailty and mortality in kidney transplant recipients. *American Journal of Transplantation* **15,** 149–154 (2015).

18. Dew, M. A. *et al.* Depression and anxiety as risk factors for morbidity and mortality after organ transplantation: a systematic review and meta-analysis. *Transplantation* **100,** 988 (2015).

19. Legendre, C., Canaud, G. & Martinez, F. Factors influencing long-term outcome after kidney transplantation. *Transplant International* **27,** 19–27 (2014).

20. Su, X., Zenios, S. A., Chakkera, H., Milford, E. L. & Chertow, G. M. Diminishing significance of HLA matching in kidney transplantation. *American Journal of Transplantation* **4,** 1501–1508 (2004).

21. Van der Vliet, J. A. & Warlé, M. C. The need to reduce cold ischemia time in kidney transplantation. *Current Opinion in Organ Transplantation* **18,** 174–178 (2013).

22. Debout, A. *et al.* Each additional hour of cold ischemia time significantly increases the risk of graft failure and mortality following renal transplantation. *Kidney International* **87,** 343–349 (2015).

23. Assfalg, V. *et al.* Repeated kidney re-transplantation—the Eurotransplant experience: a retrospective multicenter outcome analysis. *Transplant International* **33,** 617–631 (2020).

24. Frei, U. *et al.* Prospective age-matching in elderly kidney transplant recipients—a 5-year analysis of the Eurotransplant Senior Program. *American Journal of Transplantation* **8,** 50–57 (2008).

25. Assfalg, V. *et al.* Kidney Transplantation After Rescue Allocation—the Eurotransplant Experience: A Retrospective Multicenter Outcome Analysis. *Transplantation* **106,** 1215–1226 (2022).

26. Miller, G. *et al.* Kidney Transplantation Outcome Predictions (KTOP): A Risk Prediction Tool for Kidney Transplants from Brain-dead Deceased Donors Based on a Large European Cohort. *European Urology.* ISSN: 0302-2838 (2022).

27. Siedlecki, A., Irish, W. & Brennan, D. C. Delayed graft function in the kidney transplant. *American Journal of Transplantation* **11,** 2279–2296 (2011).

28. Gwinner, W. Renal transplant rejection markers. *World Journal of Urology* **25,** 445–455 (2007).

29. Dempster, N. J., Ceresa, C. D., Aitken, E. & Kingsmore, D. Outcomes following renal transplantation in older people: a retrospective cohort study. *BMC Geriatrics* **13,** 1–4 (2013).

30. Roodnat, J. *et al.* Cholesterol as an independent predictor of outcome after renal transplantation. *Transplantation* **69,** 1704–1710 (2000).

31. Khairallah, P. *et al.* The impact of COVID-19 on kidney transplantation and the kidney transplant recipient-One year into the pandemic. *Transplant International* **34,** 612–621 (2021).

32. Azzi, Y., Bartash, R., Scalea, J., Loarte-Campos, P. & Akalin, E. COVID-19 and solid organ transplantation: a review article. *Transplantation* **105,** 37–55 (2021).

33. Craig-Schapiro, R. *et al.* COVID-19 outcomes in patients waitlisted for kidney transplantation and kidney transplant recipients. *American Journal of Transplantation* **21,** 1576–1585 (2021).

34. Cholankeril, G. *et al.* Early impact of COVID-19 on solid organ transplantation in the United States. *Transplantation* (2020).

35. Chavarot, N. *et al.* COVID-19 severity in kidney transplant recipients is similar to nontransplant patients with similar comorbidities. *American Journal of Transplantation* **21,** 1285–1294 (2021).

36. Aubert, O. *et al.* COVID-19 pandemic and worldwide organ transplantation: a population-based study. *The Lancet Public Health* **6,** e709–e719 (2021).

37. Sohrabi, C. *et al.* World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery* **76,** 71–76 (2020).

38. Contini, C. *et al.* The novel zoonotic COVID-19 pandemic: An expected global health concern. *The Journal of Infection in Developing Countries* **14,** 254–264 (2020).

39. WHO. *Rolling updates on coronavirus disease (COVID-19)* https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen (2020).

40. Johns Hopkins University. *COVID-19 Dashboard* https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 (2020).

41. Lake, M. A. What we know so far: COVID-19 current clinical knowledge and research. *Clinical Medicine* **20,** 124 (2020).

42. Der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland, A. *Zahl der Infizierten steigt weiter* https://www.tagesschau.de/inland/coronavirus-deutschland-infektionen-103.html (2020).

43. Federal Ministry of Health. *Coronavirus SARS-CoV-2: Chronik der bisherigen Maßnahmen* https://www.bundesgesundheitsministerium.de/coronavirus/chronik-coronavirus.html (2020).

44. Der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland, A. *Der Ausbruch einer Pandemie* https://www.tagesschau.de/faktenfinder/hintergrund/corona-chronik-pandemie-101.html (2020).

45. Der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland, A. *Ein Virus verändert die Welt* https://www.tagesschau.de/faktenfinder/hintergrund/corona-chronik-pandemie-103.html (2020).

46. Der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland, A. *Ausnahmezustand als neue Normalität* https://www.tagesschau.de/faktenfinder/hintergrund/corona-chronik-pandemie-109.html (2020).

47. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* ISBN: 978-3-319-24277-4 (Springer, 2016).

48. Ram, K. & Wickham, H. *wesanderson: A Wes Anderson Palette Generator* R package version 0.3.6 (2018).

49. Kassambara, A. *ggpubr: 'ggplot2' Based Publication Ready Plots* R package version 0.3.0 (2020).

50. Robert Koch-Institut. *RKI Corona Landkreise dl-de/by-2-0* https://www.arcgis.com/sharing/rest/content/items/f10774f1c63e40168479a1feb6c7ca74/data.

51. R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (Vienna, Austria, 2020).

52. Robert Koch-Institut. *Coronavirus Disease 2019(COVID-19)Daily Situation Report of the Robert Koch Institute 24/06/2020* tech. rep. (Robert Koch-Institut, 2020).

53. Wei, J., Zhu, Z., Qi, Q. & Zeng, L. Patient Delay in Hospital Visiting and the Weekend Effect of Surveillance Report on Hand-Foot-and-Mouth Disease and Epidemic Parotitis in Hanzhong City, China. *Canadian Journal of Infectious Diseases and Medical Microbiology* **2020,** 11 (2020).

54. Bergman, A., Sella, Y., Agre, P. & Casadevall, A. Oscillations in US COVID-19 incidence and mortality data reflect diagnostic and reporting factors. *Msystems* **5,** e00544–20 (2020).

55. CDC COVID-19 Response Team. *Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020* tech. rep. (Centers for Disease Control and Prevention, 2020).

56. Dorjee, K. & Kim, H. Epidemiological Risk Factors Associated with Death and Severe Disease in Patients Suffering From COVID-19: A Comprehensive Systematic Review and Meta-analysis. *medRxiv* (2020).

57. Williamson, E. *et al.* OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. *medRxiv* (2020).

58. Sama, I. E. *et al.* Circulating plasma concentrations of angiotensin-converting enzyme 2 in men and women with heart failure and effects of renin-angiotensin-aldosterone inhibitors. *European Heart Journal* **41,** 1810–1817 (2020).

59. Therneau, T. M. & Grambsch, P. M. in *Modeling survival data: extending the Cox model* 39–77 (Springer, 2000).

60. Moore, D. F. *Applied survival analysis using R* (Springer, 2016).

61. Case, L. D., Kimmick, G., Paskett, E. D., Lohman, K. & Tucker, R. Interpreting measures of treatment effect in cancer clinical trials. *The Oncologist* **7,** 181–187 (2002).

62. Kaplan, E. L. & Meier, P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53,** 457–481 (1958).

63. Fine, J. P. & Gray, R. J. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* **94,** 496–509 (1999).

64. Scheike, T. H. & Zhang, M. Flexible competing risks regression modeling and goodness-of-fit. *Lifetime Data Analysis* **14,** 464–483 (2008).

65. Schwarz, G. Estimating the dimension of a model. *The Annals of Statistics* **6,** 461–464 (1978).

66. Neumair, M. *et al.* Accommodating heterogeneous missing data patterns for prostate cancer risk prediction. *arXiv preprint arXiv:2109.10141* (2021).

67. Azur, M. J., Stuart, E. A., Frangakis, C. & Leaf, P. J. Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research* **20,** 40–49 (2011).

68. White, I. R., Royston, P. & Wood, A. M. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine* **30,** 377–399 (2011).

69. Zhang, Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Annals of Translational Medicine* **4,** 30 (2016).

70. Vickers, A. J. & Elkin, E. B. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making* **26,** 565–574 (2006).

71. Vickers, A. J., Cronin, A. M., Elkin, E. B. & Gonen, M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC medical informatics and decision making* **8,** 1–17 (2008).

72. Wolbers, M., Blanche, P., Koller, M. T., Witteman, J. C. & Gerds, T. A. Concordance for prognostic models with competing risks. *Biostatistics* **15,** 526–539 (2014).

73. Gerds, T. A. & Kattan, M. W. *Medical Risk Prediction Models: With Ties to Machine Learning (1st ed.)* (Chapman and Hall/CRC, 2021).

74. Lee, E. W., Wei, L., Amato, D. A. & Leurgans, S. in *Survival analysis: state of the art* 237–247 (Springer, 1992).

75. Klein, J. P. & Moeschberger, M. L. *Survival analysis: techniques for censored and truncated data* (Springer, 2003).

76. Dellaert, F. *The expectation maximization algorithm* tech. rep. (Georgia Institute of Technology, 2002).

77. Glidden, D. V. & Vittinghoff, E. Modelling clustered survival data from multicentre clinical trials. *Statistics in Medicine* **23,** 369–388 (2004).

78. Snavely, A. C. *Multivariate Data Analysis with Applications to Cancer* PhD thesis (Harvard University, 2012).

79. Schneble, M., De Nicola, G., Kauermann, G. & Berger, U. Nowcasting fatal COVID-19 infections on a regional level in Germany. *arXiv preprint arXiv:2005.07452* (2020).

80. Wang, Q., Xie, S., Wang, Y. & Zeng, D. Survival-Convolution Models for Predicting COVID-19 Cases and Assessing Effects of Mitigation Strategies. *Frontiers in Public Health* **8,** 325 (2020).

81. IHME, C. Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European Economic Area countries. *medRxiv* (2020).

82. Mercker, M., Betzin, U. & Wilken, D. What influences COVID-19 infection rates: A statistical approach to identify promising factors applied to infection data from Germany. *medRxiv* (2020).

83. Guenther, F., Bender, A., Katz, K., Kuechenhoff, H. & Hoehle, M. Nowcasting the COVID-19 pandemic in Bavaria. *medRxiv* (2020).

84. Woody, S. *et al.* Projections for first-wave COVID-19 deaths across the US using social-distancing measures derived from mobile phones. *medrxiv* (2020).

85. Flaxman, S. *et al.* Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in European countries: technical description update. *arXiv preprint arXiv:2004.11342* (Apr. 2020).

86. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* **395,** 689–697 (2020).

87. Charu, V. *et al.* Human mobility and the spatial transmission of influenza in the United States. *PLoS Computational Biology* **13,** e1005382 (2017).

88. Mahsin, M., Deardon, R. & Brown, P. Geographically dependent individual-level models for infectious diseases transmission. *Biostatistics* **23,** 1–17 (2022).

89. Lowe, R. *et al.* The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. *Statistics in Medicine* **32,** 864–883 (2013).

90. Meyer, S., Elias, J. & Höhle, M. A space-time conditional intensity model for invasive meningococcal disease occurrence. *Biometrics* **68,** 607–616 (2012).

91. Paul, M., Held, L. & Toschke, A. M. Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine* **27,** 6250–6267 (2008).

92. Hefley, T. J., Hooten, M. B., Hanks, E. M., Russell, R. E. & Walsh, D. P. Dynamic spatio-temporal models for spatial data. *Spatial Statistics* **20,** 206–220 (2017).

93. Malesios, C., Demiris, N., Kalogeropoulos, K. & Ntzoufras, I. Bayesian epidemic models for spatially aggregated count data. *Statistics in Medicine* **36,** 3216–3230 (2017).

94. Murray, J. D. *Mathematical biology: I. An introduction* (Springer, 2007).

95. Lipsitch, M. *et al.* Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300,** 1966–1970 (2003).

96. Porta, M. *A dictionary of epidemiology* (Oxford university press, 2008).

97. Du, Z. *et al.* The serial interval of COVID-19 from publicly reported confirmed cases. *medRxiv* (2020).

98. Zhou, X., Li, Y., Li, T. & Zhang, W. Follow-up of asymptomatic patients with SARS-CoV-2 infection. *Clinical Microbiology and Infection* **26,** 957–959 (2020).

99. Nishiura, H. *et al.* Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *medRxiv* (2020).

100. Chen, X., Li, J., Xiao, C. & Yang, P. Numerical solution and parameter estimation for uncertain SIR model with application to COVID-19. *Fuzzy optimization and decision making* **20,** 189–208 (2021).

101. Postnikov, E. B. Estimation of COVID-19 dynamics “on a back-of-envelope”: Does the simplest SIR model provide quantitative parameters and predictions? *Chaos, Solitons & Fractals* **135,** 109841 (2020).

102. Truelove, S. *et al.* The potential impact of COVID-19 in refugee camps in Bangladesh and beyond: A modeling study. *PLoS Medicine* **17,** e1003144 (2020).

103. Su, L. *et al.* Evaluation of the secondary transmission pattern and epidemic prediction of COVID-19 in the four metropolitan areas of China. *Frontiers in Medicine* **7,** 171 (2020).

104. Chatterjee, K., Chatterjee, K., Kumar, A. & Shankar, S. Healthcare impact of COVID-19 epidemic in India: A stochastic mathematical model. *Medical Journal Armed Forces India* **76,** 147–155 (2020).

105. He, S., Peng, Y. & Sun, K. SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dynamics* **101,** 1667–1680 (2020).

106. Jewell, N. P., Lewnard, J. A. & Jewell, B. L. Caution warranted: using the Institute for Health Metrics and Evaluation model for predicting the course of the COVID-19 pandemic. *Annals of Internal Medicine* **173,** 226–227 (2020).

107. Markel, H. *et al.* Nonpharmaceutical interventions implemented by US cities during the 1918-1919 influenza pandemic. *Journal of the American Medical Association* **298,** 644–654 (2007).

108. Hatchett, R. J., Mecher, C. E. & Lipsitch, M. Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proceedings of the National Academy of Sciences* **104,** 7582–7587 (2007).

109. Marchant, R., Samia, N. I., Rosen, O., Tanner, M. A. & Cripps, S. Learning as We Go: An Examination of the Statistical Accuracy of COVID19 Daily Death Count Predictions. *arXiv preprint arXiv:2004.04734* (2020).

110. Atkeson, A. *et al.* On Using SIR Models to Model Disease Scenarios for COVID-19. *Quarterly Review* **41,** 1–35 (2020).

111. Schumacher, F. L., Ferreira, C. S., Prates, M. O., Lachos, A. & Lachos, V. H. A robust nonlinear mixed-effects model for COVID-19 deaths data. *arXiv preprint arXiv:2007.00848* (2020).

112. Schüttler, J., Schlickeiser, R., Schlickeiser, F. & Kröger, M. Covid-19 predictions using a Gauss model, based on data from April 2. *Physics* **2,** 197–212 (2020).

113. Goodrich, B., Gabry, J., Ali, I. & Brilleman, S. *rstanarm: Bayesian applied regression modeling via Stan.* R package version 2.19.3. 2020.

114. Millett, G. A. *et al.* Assessing differential impacts of COVID-19 on black communities. *Annals of Epidemiology* **47,** 37–44 (2020).

115. Neelon, B., Mutiso, F., Mueller, N. T., Pearce, J. L. & Benjamin-Neelon, S. E. Spatial and temporal trends in social vulnerability and COVID-19 incidence and death rates in the United States. *medRxiv* (2020).

116. Coker, E. S. *et al.* The effects of air pollution on COVID-19 related mortality in northern Italy. *Environmental and Resource Economics* **76,** 611–634 (2020).

117. Flaxman, S. *et al. Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries* tech. rep. (Imperial College London, Mar. 2020).

118. Delius, G. W. *et al.* More prevalent, less deadly? Bayesian inference of the COVID19 Infection Fatality Ratio from mortality data. *medRxiv* (2020).

119. Olney, A. M., Smith, J., Sen, S., Thomas, F. & Unwin, H. J. T. Estimating the Effect of Social Distancing Interventions on COVID-19 in the United States. *medRxiv* (2020).

120. Brauner, J. M. *et al.* The effectiveness and perceived burden of nonpharmaceutical interventions against COVID-19 transmission: a modelling study with 41 countries. *medRxiv* (2020).

121. Saez, M., Tobias, A., Varga, D. & Barceló, M. A. Effectiveness of the measures to flatten the epidemic curve of COVID-19. The case of Spain. *Science of the Total Environment* **727,** 138761 (2020).

122. Fried, R. *et al.* Retrospective Bayesian outlier detection in INGARCH series. *Statistics and Computing* **25,** 365–374 (2015).

123. Liboschik, T., Fokianos, K. & Fried, R. *tscount: An R package for analysis of count time series following generalized linear models* (Universitätsbibliothek Dortmund Dortmund, Germany, 2015).

124. Fokianos, K., Rahbek, A. & Tjøstheim, D. Poisson autoregression. *Journal of the American Statistical Association* **104,** 1430–1439 (2009).

125. Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N. The BUGS project: Evolution, critique and future directions. *Statistics in Medicine* **28,** 3049–3067 (2009).

126. Plummer, M. *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling* in *Proceedings of the 3rd international workshop on distributed statistical computing* **124** (2003), 1–10.

127. Carpenter, B. *et al.* Stan: A probabilistic programming language. *Journal of Statistical Software* **76** (2017).

128. Hoffman, M. D. & Gelman, A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* **15,** 1593–1623 (2014).

129. Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. Hybrid monte carlo. *Physics Letters B* **195,** 216–222 (1987).

130. Neal, R. M. *et al.* MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo* **2,** 2 (2011).

131. Betancourt, M. A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434* (2017).

132. Nesterov, Y. Primal-dual subgradient methods for convex problems. *Mathematical Programming* **120,** 221–259 (2009).

133. Sturtz, S., Ligges, U. & Gelman, A. R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software* **12,** 1–16 (2005).

134. Su, Y.-S. & Yajima, M. *R2jags: Using R to Run 'JAGS'* R package version 0.6-1 (2020).

135. Stan Development Team. *RStan: the R interface to Stan* R package version 2.21.2. 2020.

136. Mengersen, K. L., Robert, C. P. & Guihenneuc-Jouyaux, C. MCMC convergence diagnostics: a review. *Bayesian Statistics* **6,** 415–440 (1999).

137. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Statistical science,* 457–472 (1992).

138. Gelman, A. *et al. Bayesian Data Analysis* (Chapman and Hall/CRC, 2013).

139. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. Rank-normalization, folding, and localization: an improved R for assessing convergence of MCMC (with discussion). *Bayesian Analysis* **16,** 667–718 (2021).

140. Gelman, A., Jakulin, A., Pittau, M. G., Su, Y.-S., *et al.* A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* **2,** 1360–1383 (2008).

141. Sahai, H. & Khurshid, A. *Statistics in epidemiology: methods, techniques and applications* (CRC press, 1995).

142. Akaike, H. in *Selected papers of hirotugu akaike* 199–213 (Springer, 1998).

143. Pan, W. Akaike's information criterion in generalized estimating equations. *Biometrics* **57,** 120–125 (2001).

144. Wang, Y. *et al.* The perils of quasi-likelihood information criteria. *Stat* **4,** 246–254 (2015).

145. Hyndman, R. J. & Koehler, A. B. Another look at measures of forecast accuracy. *International Journal of Forecasting* **22,** 679–688 (2006).

146. Raynaud, M. *et al.* Dynamic prediction of renal survival among deeply phenotyped kidney transplant recipients using artificial intelligence: an observational, international, multicohort study. *The Lancet Digital Health* **3,** e795–e805 (2021).

147. Federal Statistical Office. *Genesis-Online, Dataset 12411-0018, Datenlizenz by-2-0* https://www-genesis.destatis.de/genesis//online?operation=table&code=12411-0018&bypass=true&levelindex=0&levelid=1603894527297#abreadcrumb (2020).

148. Federal Statistical Office. *Genesis-Online, Dataset 11111-0002, Datenlizenz by-2-0* https://www-genesis.destatis.de/genesis//online?operation=table&code=11111-0002&bypass=true&levelindex=0&levelid=1603894959422#abreadcrumb (2020).

149. Robert Koch-Institut. *COVID-19-Impfungen in Deutschland* https://github.com/robert-koch-institut/COVID-19-Impfungen_in_Deutschland/ (2022).

150. Robert Koch-Institut. *RKI Corona Landkreise, dl-de/by-2-0* https://www.arcgis.com/sharing/rest/content/items/f10774f1c63e40168479a1feb6c7ca74/data.

151. GovData. *DL-DE->BY-2.0* https://www.govdata.de/dl-de/by-2-0 (2022).

152. Google LLC. *COVID-19 Community Mobility Report* https://www.google.com/covid19/mobility?hl=de (2022).

153. Bundespresseamt. *Besprechung der Bundeskanzlerin mit den Regierungschefinnen und Regierungschefs der Länder vom 22.03.2020* https://www.bundesregierung.de/breg-de/themen/coronavirus/besprechung-der-bundeskanzlerin-mit-den-regierungschefinnen-und-regierungschefs-der-laender-vom-22-03-2020-1733248 (2021).

154. Bundespresseamt. *Telefonkonferenz der Bundeskanzlerin mit den Regierungschefinnen und Regierungschefs der Länder am 13. Dezember 2020* https://www.bundesregierung.de/resource/blob/997532/1827366/69441fb68435a7199b3d3a89bff2c0e6/2020-12-13-beschluss-mpk-data.pdf?download=1 (2021).

155. Bundespresseamt. *Telefonschaltkonferenz der Bundeskanzlerin mit den Regierungschefinnen und Regierungschefs der Länder am 30. April 2020* https://www.bundesregierung.de/resource/blob/975226/1749804/353e4b4c77a4d9a724347ccb688d3558/2020-04-30-beschluss-bund-laender-data.pdf?download=1 (2021).

156. Bundespresseamt. *Videokonferenz der Bundeskanzlerin mit den Regierungschefinnen und Regierungschefs der Länder am 28. Oktober 2020* https://www.bundesregierung.de/breg-de/aktuelles/videokonferenz-der-bundeskanz

lerin-mit-den-regierungschefinnen-und-regierungschefs-der-laender-am-28-oktober-2020-1805248 (2021).

157. Bundespresseamt. *Videoschaltkonferenz der Bundeskanzlerin mit den Regierungschefinnen und Regierungschefs der Länder am 25. November 2020* https://www.bundesregierung.de/resource/blob/975226/1820174/fd9794fa8b8e0ec555f005677509c242/2020-11-25-mpk-beschluss-data.pdf?download=1 (2021).

158. Stabiszewski, A. & Ebert, F. *OpenDataLab, GeoJSON Utilities, geographic data adapted from Geodatenzentrum © GeoBasis-DE / BKG 2018 VG250 31.12., status from 31.12.2018* https://www-genesis.destatis.de/genesis//online?operation=table&code=11111-0002&bypass=true&levelindex=0&levelid=1603894959422#abreadcrumb (2020).

159. DWD Climate Data Center. *Recent daily station observations (temperature, pressure, precipitation, sunshine duration, etc.) for Germany, quality control not completed yet, version recent* https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/kl/recent/ (2021).

160. Pebesma, E. J. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* **30,** 683–691 (2004).

161. Gräler, B., Pebesma, E. & Heuvelink, G. Spatio-Temporal Interpolation using gstat. *The R Journal* **8,** 204–218 (2016).

162. Wiemken, T. L. *et al.* Impact of temperature relative humidity and absolute humidity on the incidence of hospitalizations for lower respiratory tract infections due to influenza, rhinovirus, and respiratory syncytial virus: Results from community-acquired pneumonia organization (CAPO) international cohort study. *The University of Louisville Journal of Respiratory Infections* **1,** 7 (2017).

163. Quesada, J. *et al.* Incubation period of COVID-19: A systematic review and meta-analysis. *Revista Clinica Española (English Edition)* **221,** 109–117 (2021).

164. Yuan, J. *et al.* Association between meteorological factors and daily new cases of COVID-19 in 188 countries: A time series analysis. *Science of the Total Environment* **780,** 146538 (2021).

165. Islam, A. R. M. *et al.* Are meteorological factors enhancing COVID-19 transmission in Bangladesh? Novel findings from a compound Poisson generalized linear modeling approach. *Environmental Science and Pollution Research* **28,** 11245–11258 (2021).

166. Agosto, A. & Giudici, P. A Poisson autoregressive model to understand COVID-19 contagion dynamics. *Risks* **8,** 77 (2020).

167. Du, Z. *et al.* Shorter serial intervals and incubation periods in SARS-CoV-2 variants than the SARS-CoV-2 ancestral strain. *Journal of Travel Medicine,* taac052 (2022).

168. Baden, L. R. *et al.* Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *New England Journal of Medicine* **384,** 403–416 (2020).

169. Puranik, A. *et al.* Comparison of two highly-effective mRNA vaccines for COVID-19 during periods of Alpha and Delta variant prevalence. *medRxiv* (2021).

170. Cavanaugh, A. M., Spicer, K. B., Thoroughman, D., Glick, C. & Winter, K. Reduced risk of reinfection with SARS-CoV-2 after COVID-19 vaccination-Kentucky, May–June 2021. *Morbidity and Mortality Weekly Report* **70,** 1081 (2021).

171. Lundberg, A. L. *et al.* Has Omicron changed the evolution of the pandemic? *JMIR Public Health and Surveillance* **8,** e35763 (2022).

172. Veneti, L. *et al.* Reduced risk of hospitalisation among reported COVID-19 cases infected with the SARS-CoV-2 Omicron BA. 1 variant compared with the Delta variant, Norway, December 2021 to January 2022. *Eurosurveillance* **27,** 2200077 (2022).

173. Amsalu, E. *et al.* Spatial-temporal analysis of tuberculosis in the geriatric population of China: An analysis based on the Bayesian conditional autoregressive model. *Archives of Gerontology and Geriatrics* **83,** 328–337 (2019).

174. Sunengsih, N. & Jaya, I. A Bayesian spatiotemporal autoregressive model with constant and temporally varying coefficients: an application to hotspot detection of childhood tuberculosis in Bandung City, Indonesia. *Communications in Mathematical Biology and Neuroscience* **2021,** 58 (2021).

175. Wang, Z. *et al. Dynamic poisson autoregression for influenza-like-illness case count prediction* in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2015), 1285–1294.

176. Waheeb, W., Ghazali, R. & Herawan, T. Ridge polynomial neural network with error feedback for time series forecasting. *PloS ONE* **11,** e0167248 (2016).

177. Woodward, W. A., Gray, H. L. & Elliott, A. C. *Applied time series analysis with R* (CRC press, 2017).

178. Mitchell, A. & Minami, M. *The ESRI guide to GIS analysis: geographic patterns & relationships* (ESRI, Inc., 1999).

179. Elliott, P. & Wartenberg, D. Spatial epidemiology: current approaches and future challenges. *Environmental Health Perspectives* **112,** 998–1006 (2004).

180. Cressie, N. & Wikle, C. K. *Statistics for spatio-temporal data* (John Wiley & Sons, 2015).

181. Macharia, P. M. *et al.* Spatio-temporal analysis of Plasmodium falciparum prevalence to understand the past and chart the future of malaria control in Kenya. *Malaria Journal* **17,** 340 (2018).

182. Tian, L. *et al.* Spatio-temporal analysis of the relationship between meteorological factors and hand-foot-mouth disease in Beijing, China. *BMC Infectious Diseases* **18,** 158 (2018).

183. Bennett, J. E. *et al.* Particulate matter air pollution and national and county life expectancy loss in the USA: A spatiotemporal analysis. *PLoS Medicine* **16,** e1002856 (2019).

184. Cappello, C., De Iaco, S., Maggio, S. & Posa, D. Modeling ocean currents through complex random fields indexed in time. *Mathematical Geosciences* **53,** 999–1025 (2021).

185. Bai, Y.-t., Jin, X.-b., Wang, X.-y., Wang, X.-k. & Xu, J.-p. Dynamic correlation analysis method of air pollutants in spatio-temporal analysis. *International Journal of Environmental Research and Public Health* **17,** 360 (2020).

186. Hepler, S. A. & Erhardt, R. J. A spatiotemporal model for multivariate occupancy data. *Environmetrics* **32,** e2657 (2021).

187. Severiano, C. A., e Silva, P. C. d. L., Cohen, M. W. & Guimarães, F. G. Evolving fuzzy time series for spatio-temporal forecasting in renewable energy systems. *Renewable Energy* **171,** 764–783 (2021).

188. Gao, N. *et al.* Generative adversarial networks for spatio-temporal data: A survey. *Association for Computing Machinery Transactions on Intelligent Systems and Technology* **13,** 1–25 (2022).

189. George, E. I. The variable selection problem. *Journal of the American Statistical Association* **95,** 1304–1308 (2000).

190. Walter, S. & Tiemeier, H. Variable selection: current practice in epidemiological studies. *European Journal of Epidemiology* **24,** 733–736 (2009).

191. Heinze, G. & Dunkler, D. Five myths about variable selection. *Transplant International* **30,** 6–10 (2017).

192. Robert Koch-Institut. *Robert Koch-Institut: COVID-19-Dashboard* `https://experience.arcgis.com/experience/478220a4c454480e823b17327b2bf1d4/page/Landkreise/` (2022).

193. World Health Organization. *WHO Coronavirus (COVID-19) Dashboard* `https://covid19.who.int/` (2022).

194. Maryland Department of Health. *Maryland COVID-19 Data Dashboard* `https://coronavirus.maryland.gov/` (2022).

195. Johns Hopkins University. *COVID-19 Dashboard* `https://coronavirus.jhu.edu/map.html` (2022).

196. Cheng, C., Barceló, J., Hartnett, A. S., Kubinec, R. & Messerschmidt, L. COVID-19 government response event dataset (CoronaNet v. 1.0). *Nature Human Behaviour* **4,** 756–768 (2020).