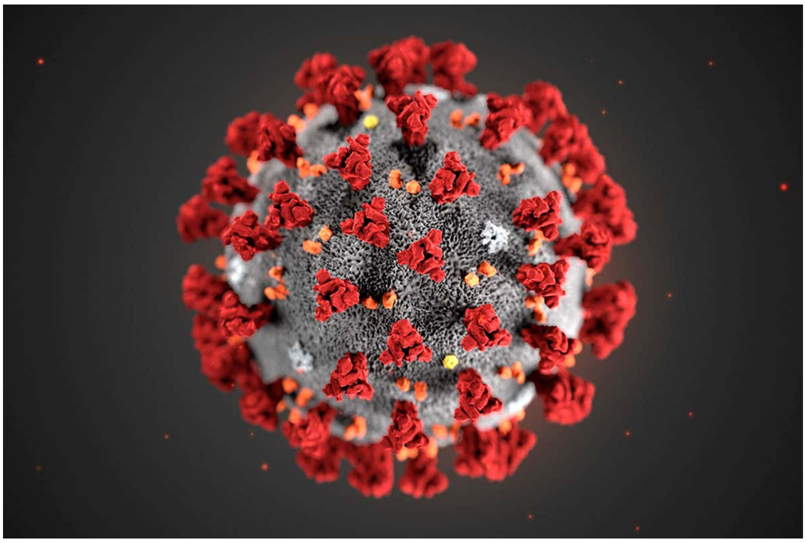
Citadel Europe regional datathon

Non-compliance: Following the path of least resistance in Covid-19 policy formulation.



Team 1: Keelan McMahon, Charlotte Roman, Stefan Stein and Patrick Zietkiewicz

1 Executive Summary

Covid-19 Modelling

It is vitally important for a country to have an accurate idea of how case numbers, hospitalizations and deaths will change over time. The number of hospitalizations and deaths depends on the number of cases and, the number of cases in turn depends on the reproduction number R(t). For this reason an accurate measure of R(t) is essential for any analysis of this data, we use a susceptible-exposed-infected-removed (SEIR) epidemiological model to attempt to give better approximations than the ones provided in the OWID dataset.

Clustering

We wanted to give a broad view of Europe, however, we understand that there is much heterogeneity amongst these countries. To address this, we decided to group countries with respect to variables that are relevant to how a country may cope with Covid-19. Subsequently, we clustered the countries based on age group susceptibility, wealth (via Gross Domestic product per capita and the GINI index) and the time that a country had to prepare from the first reported case of Covid in Europe (a "warning" time).

Prediction

We use time series modelling to forecast the ICU occupancy in a particular set of countries using the lagged number of cases. We also showed that the an approximation of the number of deaths in a two weeks time can be reasonably predicted by the case positivity rate of today. The purpose of this model is to serve as a tool for policymakers by giving them an idea of what ICU occupancy could be in 10 days time. This, coupled with other relevant information would allow them to make more informed decisions about restricting the movement of citizens.

Non-Compliance and Behavioural Fatigue

Here, we introduce some sources that discuss the impact of non-compliance and behavioural fatigue on policymaker's decisions.

Mobility

Mobility has a key impact on the spread of disease since it indicates how travel behaviours were impacted by the pandemic. We explore the trends in mobility data in relation to the reproduction number. The impact of regulation on mobility is also explored with case examples studied to analyze how response measures have impacted behaviour. Finally, we compile the reproduction numbers we found from our SEIR model with the mobility data and stringency index to understand how regulation compliance impacted the reproduction rate.

Shiny App

We have developed an app, please click here, to accompany our investigation. It allows users to better understand parts of our analysis in an interactive manner.

Research Question

This leads us to the following research questions:

- How can we reliably predict how policies will effect the number of Covid cases, and in particular Covid cases that are resulting in hospitalizations and deaths?
- Can we find evidence of increased non-compliance with government measures as restrictions go on, and what type of restrictions lead to non-compliance?

2 Report

2.1 Introduction

The virus known as the 2019 Novel Coronavirus (or Covid-19), has spread rapidly across the world since it was first detected in December of 2019. It caused the ongoing global pandemic that has had devastating impact on patients across the world. As we are still in the very early stages of the disease very little is known about how to most effectively treat patients with the virus, and primary treatment is currently just symptomatic. As a result of this, policymakers are having to rely on non-pharmaceutical interventions to suppress the spread of the disease. There have been studies showing that these early interventions have been effective in saving lives and suppressing the virus (Balmford, Annan, Hargreaves, Altoè, and Bateman, 2020). However, the reintroduction of the same interventions might not be as effective in the future as citizens may be less willing to comply. There are a number of possible reasons for increasing non-compliance, for example "adherence fatigue" may start if people feel that they have had no decision on their restricted movements. People may also not be compliant for other reasons such as economic necessity (Michie, West, and Harvey, 2020) or suffer from something equivalent to evacuation fatigue that has been reported for a variety of disaster types such as wildfires, hurricanes and mud-slides (Harvey, 2020). Non-compliance would completely undermine the effectiveness of any intervention.

One relevant concept from behavioural economics is nudge theory, which proposes positive reinforcement and indirect suggestions as ways to influence group decision-making. The term "nudge" was popularised in the context of health decisions by Thaler and Sunstein (2009), who define it as "any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentive". Some nations, such as the United Kingdom, the Netherlands, and Sweden, were known to consider strategies using behavioural nudges. The UK Prime Minister, Boris Johnson, was well-known in the early stages of the pandemic to suggest "herd immunity" as a strategy[1]. Similarly, in the Netherlands strategy of Prime Minister Rutte was announced that to be "maximum control" of the country, rather than enforce strict lockdowns[2]. Although both of these countries later enforced restrictions due to the growing pressure from scientists, the question remains as to how effective "nudges" are at changing the behaviours of the population. Using behavioural economics to fight the spread of coronavirus would have been an extreme risk at the beginning of the pandemic when limited information was known about it. Nevertheless, nudge theory and other techniques from behavioural economics could play a key role in reaching a pre-pandemic-like society.

In this paper, we looked at the effectiveness of government policies in suppression of the virus. As has been done in Balmford et al. (2020), we measure the strength of a government's restrcitions by the reported *stringency_index* found in the `owdi/covid_date.csv` file. We model Covid-19 cases in a country using the susceptible-exposed-infected-removed (SEIR) epidemiological model introduced to model Covid-19 by the Irish Epidemiology Modelling Advisory group. We justify its inclusion by showing that the model via which the provided $R(t)$ values were calculated may be causing inaccuracies. The full details of the SEIR model are provided in section 2.2. A description of the model is seen in Figure 2 and it is described in detail in IEMAG (2020). We will also model and predict hospitalizations and deaths based on a time series analysis involving case data. And compare how well countries have done in suppressing and dealing with the virus.

Our contribution to the literature is on extending this comparison to investigate whether increasing non-compliance is evident in 28 European countries specified in table 1 and to see if it varies by the stringency of the government enforced lockdown. We do this by using Google mobility data (https://www.google.com/covid19/mobility/), and use this as a measure of policy adherence. This mobility data has been shown to predict the spread of the disease in some regions Wang and Yamamoto (2020). We will see how mobility has developed over the course of the pandemic and if it leads to additional reported cases. The discussion can be found in Sections 2.6 and 2.7.

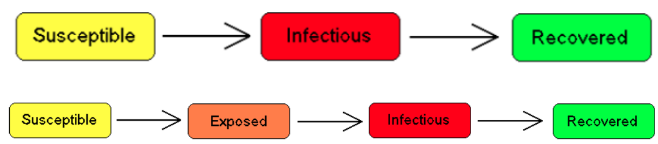


Figure 1: The most simple way to model infections is by the SIR model (top). It considers the population dynamics of Susceptible Infectious Recovered. In the SEIR model (bottom), the additional stage is introduced to account for people that have been exposed to the virus but are not yet infectious. (Image source: wikipedia.com)

We also recognize that Covid cases differ greatly in their severity. Many people catching Covid survive the virus without complications nor adverse effects, while others become seriously ill, needing treatment in hospitals, or may even die. This suggests, that analysis of the case numbers as such, even when normalizing them by the number of tests conducted, is the wrong measure to inform policy. In other words, a policymaker whose primary objective is protecting their population as best as possible, should not focus on reducing the number of cases per se, but rather on reducing the number of severe cases of Covid. Clearly, the two are strongly correlated and our second research question is motivated by the desire to understand how exactly ICU cases are driven by case numbers and other covariates. In particular, we wanted to test the assumptions of some models which assume that hospitalisation cases are just, some proportion (between 20-30%) of Covid-19 cases and that ICU in turn is between 15 % and 20 % of hospitalization, and deaths in turn occur as a fixed proportion of these variables, we will attempt to see if any more information may be useful in predicting hospitalisation numbers and deaths. An in depth analysis is provided in Section 2.4.

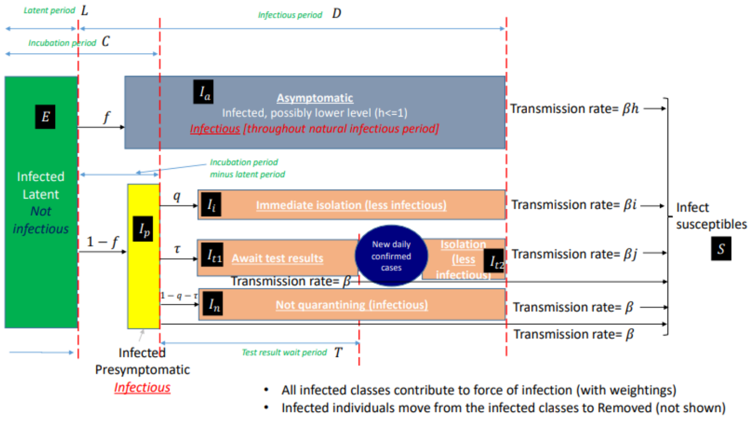


Figure 2: Model compartment structure

2.2 Data Cleaning and the SEIR model

Before attempting any analysis of the data we went through a process of correcting corrupt or possibly inaccurate data in the provided Owid dataset. The first thing we did was look at the four variables that we thought would be of most interest for each country; daily Covid cases, daily totals of number of covid patients in hospital and in intensive care units (ICU), and also daily Covid deaths. In some countries there were some recorded dates with no associated daily cases. If these dates were before the first case day, we removed these rows from our analysis. Another thing we noticed was that some Covid cases were negative. To correct for this, we assumed these were adjustments to the cumulative total and were in response to previous over-reporting. We adjusted the values from previous days at a rate proportional to their suspected over-reporting, from comparing the weekly totals in the Owid data with the weekly total totals in the provided ECDC data . With some of the hospitalization variables, ICU variables and death variables, some of the countries did not report them. In this case we used the hospitalisation, intensive care admission, and mortality rates estimates, based on ECDC rapid risk assessments on experience of the pandemic, described in Figure 3.

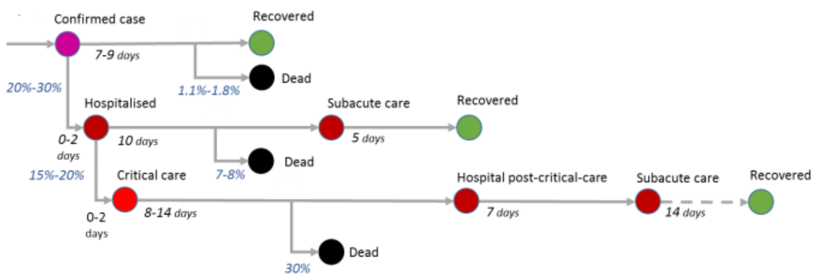


Figure 3: Model of disease and care: The parameters here are clinical, and used to estimate healthcare demand from the outputs of the SEIR model. The hospitalisation, intensive care admission, and mortality rates are the estimates, based on ECDC rapid risk assessments and early experience of the pandemic.

We also noted that we had been provided with effective reproduction, or R, numbers, a dynamic estimate of the average number of secondary cases generated by a single typical case over the course of its infectious period, this is a very important statistic because if its value is greater than 1, the number of infected people grows exponentially. Therefore if this was an accurate measure of R it would be hugely beneficial of our analysis of how well a country has responded to the outbreak of the virus. So we looked at the model and data that went into generating it (Arroyo-Marioli, Bullano, Kucinskas, and Rondón-Moreno, 2021) and found that the calculation of the $R(t)$ value was done using the simplified SIR model, which does not take into account the full complexity of the development of Covid-19. It does not take into account an incubation period, and it does not take into account asymptomatic cases. While these are difficult to measure, the literature provides various ranges for them, suggesting that they account for between 18 and 82 percent of cases (McAloon, Collins, Hunt, Barber, Byrne, Butler, Casey, Griffin, Lane, McEvoy, Wall, Green, O'Grady, and More, 2020).

This immediately led us to our first finding: The $R(t)$ value as reported in `owid_covid_data.csv` is **not suitable** for capturing the actual spreading behaviour of the pandemic. Our first contribution is the introduction of an SEIR model which would give a much better approximation of the R value, upon which we base the rest of our analysis. To do this we used country-level SEIR models which like the SIR model assumes fully-mixed populations within the country. Despite this simplification, SEIR models provide a useful tool for scenario-analysis. At each time-step, every individual is assumed to be either susceptible, exposed, infected or recovered, and move between the components in a direct acyclical way. There are reports of recovered people being susceptible again but we ignored them in these calculations. We recognize that this leaves scope to extend this model to include this if future evidence warrants their inclusion. Information on the time taken to move between components has been studied extensively which provides useful information to inform the model parameters. Another reason we decided against using the SIR R value was because it was computed using a different dataset of confirmed cases than the one we were provided: it uses datasets compiled by John Hopkins University.

A full description of the SEIR model that we based our analysis, can be found in IEMAG (2020). The SEIR model described in figure 2 was used. The dynamics evolve according to the differential equations in IEMAG (2020) (1.2), and they also describe likely priors for some of the model parameters. The daily Covid-19 cases at day t, are modelled by a negative binomial random variable. To model the mean parameter of the negative binomial distribution, we use a thin-plate regression spline. We used 10 thin-plate basis functions to achieve a satisfactory level of fit, outlined in the paper. The resulting model is a negative binomial generalized additive model (GAM). To account for parameter uncertainty in modelling cases, the model was fitted in a Bayesian framework to the provided owid data using the `brms` R package (Bürkner, 2017). The `brms` R package interfaces with Stan to generate samples from the posterior distribution for the model parameters. We have calculated time-developing credible intervals associated with each of the dynamic parameters, but we focused on medians of the sampled parameters and specifically R in our analysis and ignored the uncertainty in the modelling parameters as we were time restricted. We point out, though, that the way in which the code was written by them, makes scenario analysis possible, as has been done in IEMAG (2020), who show that lowering R values below 1 is the driving factor for disease suppression. We will build on this work by investigating factors behind $R(t)$ value dynamics and not assume they can be automatically adjusted with policy interventions.

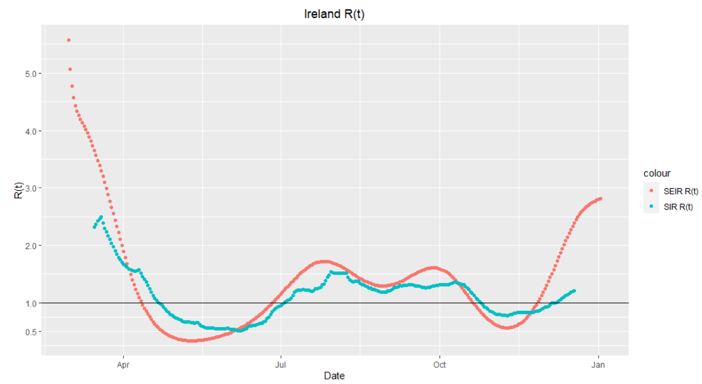


Figure 4: Illustration of discrepancies between the SEIR and SIR generated R(t) values.

Comparisons of the SEIR and SIR values showed some discrepancies, Figure 4, even when ignoring the initial values, which just reflect that in the early stages of the pandemic, neither the SEIR model and SIR model accuratley describe the spread of Covid-19. These discrepancies appear be driven by the assumption in the SIR model that there is no latent period, which justifies us using SEIR $R(t)$ for our analysis. Without accounting for this, the SIR $R(t)$ may have an erroneous lag. This is especially evident when looking at the points at which the $R(t)$ estimate cross one, every time the SIR overestimates this crossing by a value that seems to approximate likely values of this incubation period from the literature (3.9 days to 5.9 days). Because of how important approximating when $R(t)$ crosses this threshold, we will use SEIR $R(t)$ for our analysis.

Having calculated these R values we were now ready to do our analysis.

2.3 Clustering

We recognize that the course of the pandemic differs greatly between countries and is dependant on country specific variables. Most importantly, it has been established that the severity of Covid significantly depends on age (Balmford et al., 2020) and, all other things considered, older sufferers are more likely to die from contracting Covid-19 than are younger people (Dowd, Andriano, Brazel, Rotondi, Block, Ding, Liu, and Mills, 2020). This means that it would not be a fair comparison to compare countries with significantly different age distributions. Also, age groups have different levels of susceptibility to catching the disease, as suggested by Figure 5, which shows that the age group of under 15s is disproportionately less susceptible to catching Covid, whereas those aged 80 and older, who are disproportionately more susceptible to catching it. Therefore, we re-weighted the age distribution according to this susceptibility.

Health outcomes might also differ because of within-country variation in wealth (Marmot, 2005). In general, richer countries are better positioned to fight the spread of a pandemic, which is why we decided to include the latest GINI index (`https://data.worldbank.org/indicator/SI.POV.GINI`), as well as per capita GDP (`https://data.worldbank.org/indicator/NY.GDP.PCAP.CD`) as per current Worldbank data, as covariates for our clustering.

Finally, previous studies (Fraser, Riley, Anderson, and Ferguson, 2004, e.g.) have highlighted that early detection may play a crucial role in halting virus spread. Hence, it seems plausible that countries which were exposed to Covid-19 earlier in the pandemic, therefore had less time to prepare, faced worse consequences. To account for this, we included "warning time" as a variable in our clustering, which we measured as follows. Out of the 28 countries we study, we found the date of the first recorded case of Covid in any of these countries, the "day zero", so to speak (e.g. 24^{th} of January, France). The warning time for each other country was the time difference in days between the first recorded Covid case in that country and day zero.

Country list			
Austria	Belgium	Croatia	Cyprus
Czechia	Denmark	Estonia	Finland
France	Germany	Greece	Hungary
Iceland	Ireland	Italy	Latvia
Lithuania	Luxembourg	Malta	Netherlands
Norway	Poland	Portugal	Romania
Slovakia	Slovenia	Spain	Sweden

Table 1: List of countries we decided to study.

We hypothesise that countries within the same cluster have had similar starting conditions when the pandemic broke out and in our the subsequent analysis, we will be comparing performance of countries within each of their respective clusters, to ensure a fair comparison.

In detail, we

1. Focus on countries for which we have access to the age distribution, which in our case are the countries listed in `2_ecdc/agerangenotificationeu.csv`. The countries under consideration are listed in Table 1.

2. Cluster these countries based on the variables outlined above using `kmeans` clustering. Specifically, the variables we used were: the percentages of the population that is aged below 15, aged $15-24$, aged $25-49$, aged $50-64$, aged $65-79$ and aged 80, the latest GINI index (`https://data.worldbank.org/indicator/SI.POV.GINI`) as well as per capita GDP (`https://data.worldbank.org/indicator/NY.GDP.PCAP.CD`) as per current Worldbank data. We also computed the warning time a country had to prepare for the outbreak of Covid. It is also known that some age groups are disproportionately susceptible to catching Covid, as suggested by figure 5. We therefore re-weighted the age distribution according to this susceptibility. We selected the optimal number of clusters

using the “elbow method”. The corresponding elbow plot and the resulting clusters are shown in figure 6.

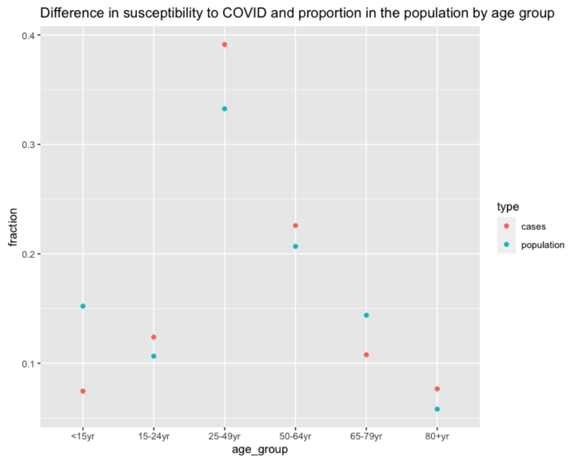


Figure 5: Fraction of age groups in the entire population of the 28 counties under study (blue), and fraction of that age group in the number of reported Covid cases. This figure shows, for example, that the age group of less than 15-year olds catches Covid more rarely than one might expect if susceptibility was evenly distributed amongst all age groups, whereas those aged 80 and older catch it disproportionately more frequently.

We conclude that countries within the same cluster have had similar starting conditions when the pandemic broke out. In our the subsequent analysis, we will be comparing performance of countries within each of their respective clusters, to ensure a fair comparison.

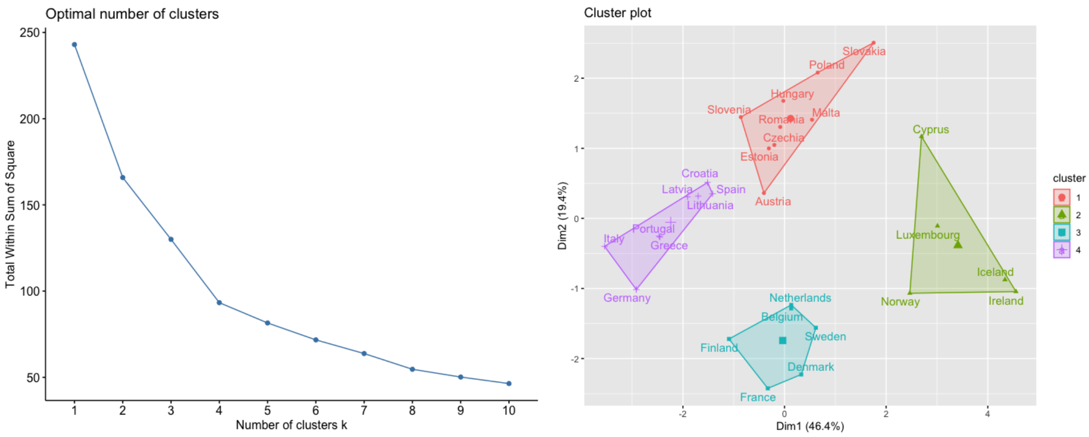


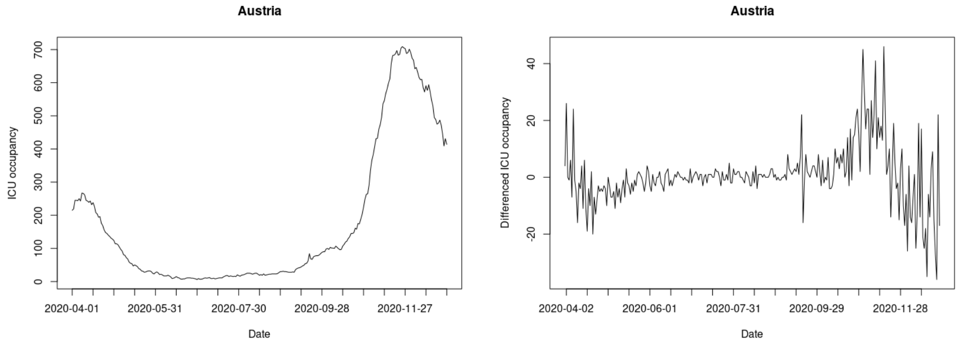
Figure 6: Elbow plot and resulting clusters. The elbow plot on the left suggests using four clusters, which are shown on the right.

2.4 Time Series Forecasting

2.4.1 Approach

As discussed previously, ICU occupancy is a critical measure that can direct policymakers and government. Thus, it is important to have a gauge of what ICU occupancy will be in the near future. To address this, we propose using a time series modelling procedure to best predict ICU occupancy. We have deemed a 10-day forecast to be reasonable enough as it gives the policymakers time to tighten or relax any lockdown measures.

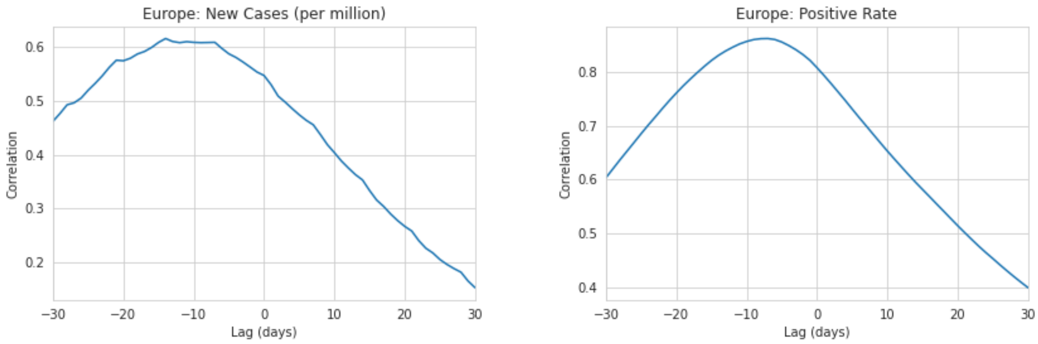
Figure 7a shows the ICU occupancy in Austria over time: trend, seasonality and stochasticity are all clearly visible. The two peaks present a problem, in a modelling sense, as they are infrequent (occurring only twice) and so it is not feasible to accurately estimate this seasonality component. The differenced time series in Figure 7b offers some hope in modelling as the trend is effectively gone, although, there are pockets of high variance at the start and near the end of the series.



(a) **(b)**

Figure 7: (Left) Austria ICU occupancy and (right) differenced Austria ICU occupancy.

There are obvious warning signs for when ICU occupancy may increase or decrease, for example, increased positive Covid tests. A time series model which only considers the time series itself is ignorant of these potentially illuminating variables. In light of this, we decided to include external regressors in the model in an attempt to incorporate the information present in increasing case numbers and other variables. This was done by lagging variables and plotting them against ICU occupancy hoping to find evidence of a correlation. After doing so, we deemed the variables new cases per million and positive rate (which is the proportion of Covid tests that are positive) as being the most significant. Evidence of this is seen in Figures 8a and 8b below, which is averaged over all European countries. There is a clear correlation between ICU occupancy and both of these variables when they are lagged by 10 days, i.e. the new cases per million today would inform the ICU occupancy 10 days from now.

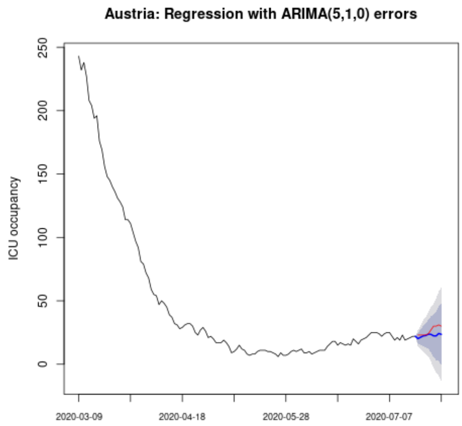


(a) **(b)**

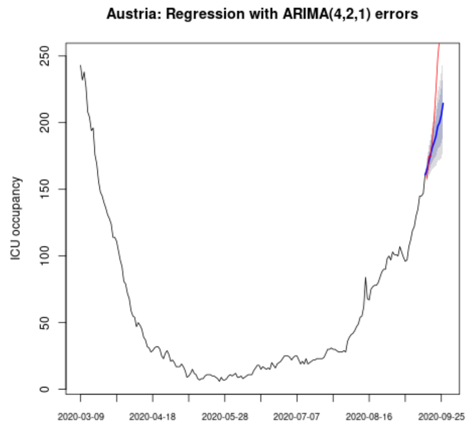
Figure 8: Lagged correlation for new cases per million and positive rate averaged over all the countries.

We opted for a simpler time series model, an autoregressive integrated moving average, known as an ARIMA(p, d, q) where p, d, and q are the number of autoregressive components, how many times the series is differenced and the number of moving average components respectively. Now it would not be reasonable to have the exact same ARIMA model as data comes in, but rather, to update our estimation and model choice. To do this, we used the `auto.arima()` function from the `forecast` package (Hyndman and Khandakar, 2008) in `R` (R Core Team, 2020). We pre-specify the maximum number of autoregressive and moving average components, all subsequent models are then compared via the Akaike Information Criterion (AIC) resulting in a *best* model. Additionally, the `auto.arima()` function tests the time series for independence and subsequently differences it if not independent.

In order to assess the accuracy of our approach, we started the models roughly halfway through the available data to see if the predictions captured what actually happened, and whether they would have been useful at the time. Figure 9 shows two examples of our model's predictions for Austria in two different time periods. Figure 9a shows how the model performs when ICU cases are relatively low and Figure 9b shows a forecast when cases are rising. The light blue shading indicates a 95% confidence interval and the darker shade a 80% confidence interval for the predicted ICU cases (blue line). An entire run through for Austria, and other countries, is visible in the Covid predictor App. The model performs reasonably well as the confidence bands capture the actual ICU occupancy (red line) most of the time, however, the forecast can be poor when there is a sudden change in ICU occupancy trajectory such as when cases start rising suddenly, or there is a peak. After a few days, the model begins to *catch up* and once again give reasonable predictions.



(a) Data up to 27/07/2020.



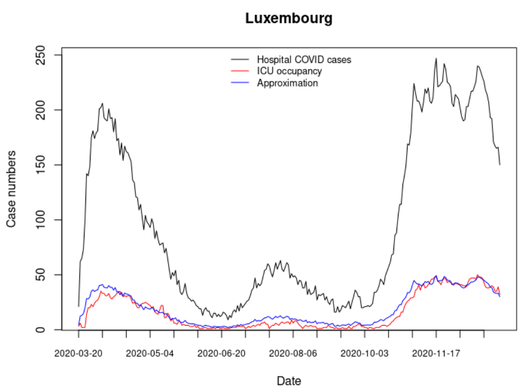
(b) Data up to 26/09/2020.

Figure 9: Illustration of ARIMA models for two different time periods in Austria. The black lines are actual ICU cases. The blue line is the prediction of the ARIMA model and the red line are the actual future cases. Dark shaded region is a 80% confidence interval. Light shaded region is a 95% confidence interval.

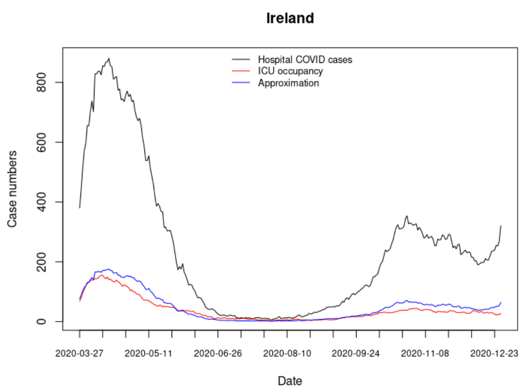
2.4.2 Dealing with Missing Data

Unfortunately, the owid data contained a lot of missing values. When there was no value present for ICU occupancy and it clearly was non-zero, we simply removed it from the data. This method is reasonable as the missing data, if present, occurred either at the beginning or towards the end of the date range. Thus, we still had the bulk of the data to deal with.

In some instances, there was none or an insufficient amount of ICU data. In this scenario, we used the rule of taking the number of Covid cases in hospital and multiplying it by 0.2 as a crude estimate for ICU occupancy, based on Figure 3 where the estimated proportion of hospitalised patients going into critical care is $0.15 - 0.2$. Figures 10a and 10b give evidence to suggest that this is an acceptable approximation. Greece had no hospital or ICU data and so was excluded from the analysis. Additionally, Lithuania had insufficient observations to fit a model and so it was excluded too.



(a) Luxembourg



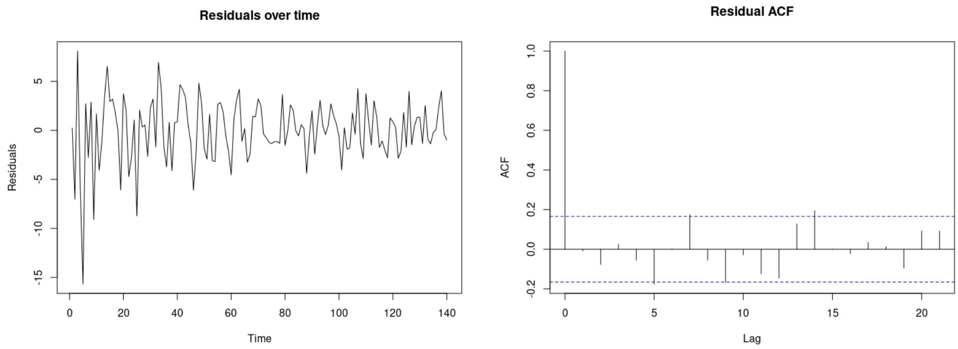
(b) Ireland

Figure 10: Crude estimates of ICU occupancy.

2.4.3 Residual Analysis and Discussion

An analysis of how well the ARIMA model for Austria on data up until and including the 26$^{\text{th}}$ July 2020 is fitting is presented below, see Figure 7a for the original data. This serves as an example, as one could perform this analysis for other fitted models. The main aspects to check are that the residuals are mean zero, constant variance and that they are independent. The resultant model is an ARIMA(5,1,0) with new cases per million and positive rate as external regressors.

Figure 11a shows that the residuals are centred around zero. However, the variance has a wider spread near the start of the time period which gradually decreases with time. So, there is not a strong case for homoscedastic variance. The auto-correlation plot (ACF) in Figure 12b does not show signs of significant correlation. This is backed up by performing a Ljung-Box test with R's `Box.test()` function which has the null hypothesis of the the data being independent, the p-value > 0.9 gives no evidence to reject this statement.



(a) (b)

Figure 11: Residual plot and ACF of residuals of ARIMA(5,1,0) with new cases per million and positive rate as external regressors.

Although Figure 11 does not cause immediate concern, it is important to note that this may not be true for the other countries considered in this model. It is easy to see that when the ICU case numbers increase, they do so rapidly, whilst when occupancy is low, there is less variance in the daily changes. Thus, a possible improvement to this model would be to allow for heteroscedastic variance explicitly by using a generalised autoregressive conditional heteroscedastic (GARCH) model.

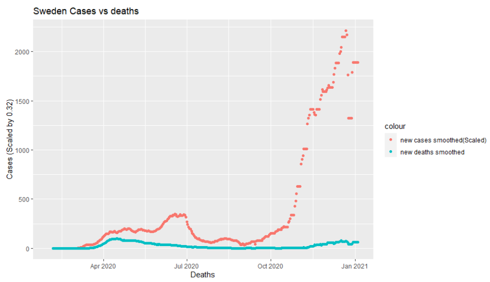
Lastly, it is important to remember what these models are ignorant of. For example, one would expect travel to occur around the Christmas holidays. This would have been the first time we would observe data in this setting and it is unlikely that the modelling approach adopted here would fully capture the increase in cases that would come from mass inter- and intra-country travel. Thus, these models should not be relied on by themselves, but rather accompanied by other forms of evidence which together give a more comprehensive outlook.

2.5 Death Forecasting

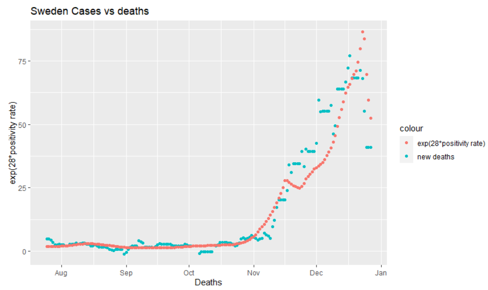
One of the variables of most concern to policymakers is the number of people who are dying from the pandemic. Previous models attempted in predicting deaths, assumed that they are just some percentage of all reported cases. This figure, for example, used by the Irish epidemiological modelling agency in the early stages of the pandemic, assumed that the proportion is between 2.3% and 4.2%. But, as evidenced in Figure 12a, we show that this fixed proportion assumption does not accurately describe the data after the initial period, and predicting case numbers may not be possible by simply using positive case data alone. Figure 12 shows that a much better approach to model future deaths is done from the current positivity rate, which takes into account both positive tests and total tests conducted. As seen in the simple plot below, deaths are reasonably well predicted by the positivity rates with a lag of fourteen days. We fit a simple model between by a fourteen day lagged positivity rate and Covid deaths and found it was reasonable at predicting deaths. This fit was consistent for all countries with R-Squared values between 0.6-0.8 for the European countries and residual analysis seem to suggest that the assumptions of the model were reasonable. With this model, we fitted a model parameter for each country, which predicts based on the case data today, the expected deaths in fourteen days time. According to the formula

$\mathbf{E}[\log(\text{deaths}(t+14)+1)] = \text{Positivityrate}(t)K.$

This fitted K value was between 1.37 and 4.1 for all of the countries, which given that number of death given positive rate, increases by an exponential rate with this variable, really shows that there was huge differences between the relationship between the number of deaths given positive rate in a county, which mainly reflected a countries testing capacity. We attempted to compare it between the clusters identified based on factors know to affect Covid. The There seemed to be some differences between the clusters, with clusters reflecting the more susceptible ccounting having much significantly larger values of K.



(a)



(b)

Figure 12: Evidence to suggest that using positivity rate instead of just new cases may be a predictor of deaths.

2.6 Non-Compliance and Behavioural Fatigue

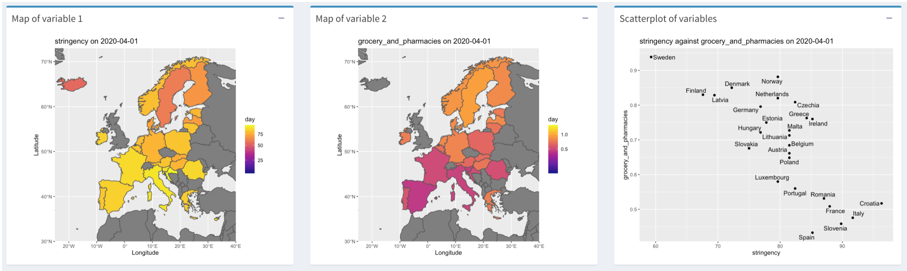
Very little has been published on measuring behavioural fatigue in response to Covid-19 regulations, and the studies that have been conducted are somewhat dismissive of the idea. Harvey (2020) even go so far as to say that because they have found no evidence of it, that behavioural fatigue is not a real phenomenon, and it must be a naïve construct or a policy contrivance. Also, in evaluating claims made against the existence of "behavioural fatigue", some studies are very restrictive of their definition of fatigue. For example, Michie et al. (2020) argue evidence that stress associated with increasing financial worries over time will undermine support for some policies, but says that this should not be considered fatigue but as "resisting hardship". They also say that an individual's concerns about the government's approach may also affect compliance but this also shouldn't be classed as fatigue. For this reason, we prefer using the term increasing non-compliance rates to avoid any ambiguity.

The popular view of progressive non-compliance or behavioural fatigue may be reflective of the fact that the idea of behavioural fatigue was espoused by many governments in their initial response to the virus. They often decided to cite behavioural fatigue as a reason not to introduce necessary lockdowns which resulted in higher levels of the disease and death within the population Balmford et al. (2020). Relying solely on behavioural concepts to suppress the disease has been ineffective, but that does not mean they should not be a consideration in policy formation. Studies that have measured non-compliance rates have relied on self-reported measures of compliance Fancourt, Bu, Mak, and Steptoe (2020). This measure relies on the participant both understanding what the current regulations are and also that the veracity of their response was not effected by priming bias in the way the questions were asked.

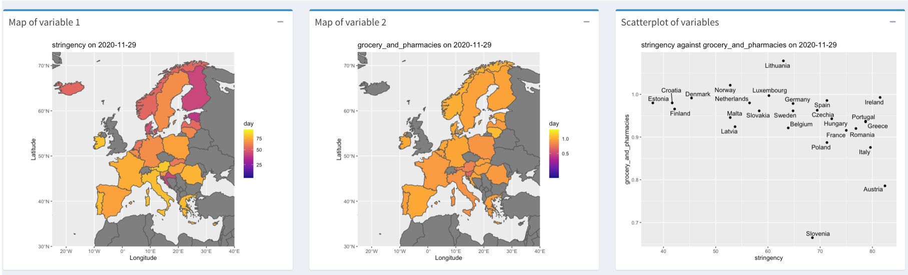
To remove these assumptions, we will look at approximating compliance rates by looking at aggregated mobility data within a country over the course of a lockdown. The data shows how visitors to categorized places change compared to baseline days. The baseline day they used is the median value from the 5-week period Jan 3rd – Feb 6th 2020. The main limitation of this aggregated data is that it is only available from the lead up to the pandemic and, therefore, it was difficult to adjust for seasonal variability which may be present in some of the mobility measures. One of their measures of park visits had a clear seasonal component, so we decided not to include it in our analysis. We also attempted to account for seasonal variability by looking at hours of sunlight within a country on a date, assuming seasonal variability in mobility may be explained by this. However, we were unable to find a reliable source for this data. If we were to extend the analysis we would take this into account but due to lack of available reliable data, we did not adjust for seasonal variability in mobility in this report.

We illustrate this behaviour in Figures 13a-13b. In early April (13a), when stringency was high across Europe, we can see that people went less to grocery stores or pharmacies as measured by the Google mobility data. On the other hand, in late November, another period of similarly high stringency, mobility remained much higher than in April, suggesting, that repressive measures may lose their effectiveness when used repeatedly. These figures are screenshot taken from our accompanying R Shiny app (`https://`

stefanstein.shinyapps.io/covid/, under the "Maps" tab) and we invite the reader to try different date values and variables themselves to see how their interaction changes over time. An in-depth analysis of this phenomenon for selected countries is presented in Section 2.7.



(a) Stringency and mobility in Europe in April.



(b) Stingency and mobility in Europe in November.

Figure 13: Stringency index compared with grocery and pharmacy mobility (Google).

2.7 Assessing Behaviour Through Mobility Data

Mobility of populations is a crucial factor in the spread of coronavirus. The amount a population is moving will affect the reproduction rate and move the virus to more geographical regions within a country. Therefore, mobility is essential in predicting the number of cases, hospitalisations, and ICU admissions. Since we were not provided with mobility data, we found mobility data from two sources, Apple (https://covid19.apple.com/mobility) and Google (https://www.google.com/covid19/mobility/). These data sources show the number of routes made using the navigation applications and can also tell us the type of journey to indicate the purpose of travel.

First, we use the Apple mobility data for exploratory data analysis. This dataset records the number of routes made using Apple Maps as a percentage compared to the baseline number made on January 13^{th} 2020. For example, a value of 200 tells us that mobility was twice as great as January 13^{th}. Thus, the data is standardised based on typical usage of the Apple Maps application. The data also tells us whether the journey was made by driving, walking, or transit (public transport services). We only consider driving and transit data as these were found to have the most correlation with the number of cases and positive rate from the Owid dataset.

Figure 14 shows the trends in driving and transit mobility. In both types of transport, there is a drastic reduction in mobility in March. In general, this then slowly increases to a peak in the late summer months. Apple mobility data is only provided from January 2020 onwards. Thus, it is difficult to determine if a peak of travel in the summer months is from seasonal trends.

However, we can compare the European mobility data with the non-European mobility data, shown in Figure 17 (Appendix). In non-European data, we do not see an overall peak of travel over the summer. This could suggest that the reduction in travel in Autumn may be correlated to Covid since Europe was one of the worst-hit regions by the pandemic. Nonetheless, there could be cultural differences or other factors that impact travel, so we only use this data to suggest not confirm relationships between mobility and Covid.

To understand the possible correlations of mobility with Covid, we plotted the driving and transit mobility against four key indicators of the impact of Covid for each country. Since there are too many countries to include, we also plotted an mean of this data which can be seen in Figures 15 and 18 (Appendix). In Figure 15, we see that increasing mobility levels in the summer months did not increase the number of cases, deaths or positive rate. This may be due to compliance with social distancing

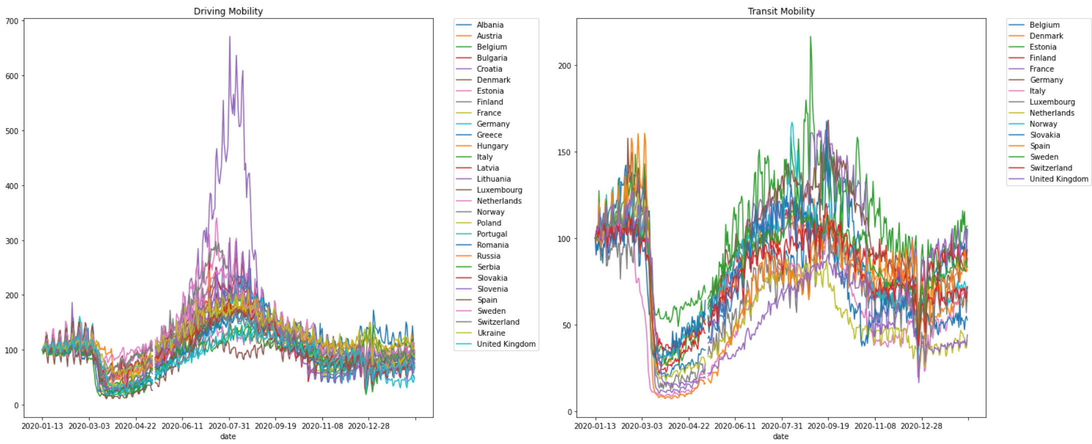


Figure 14: The Apple mobility data for European countries.

measures. The stringency and driving mobility have a negative correlation before June; as stringency is increased, mobility reduced and vice versa. Then over summer, mobility peaks and reduces despite stringency levels staying relatively constant. Similar effects occur with transit data (see Appendix Figure

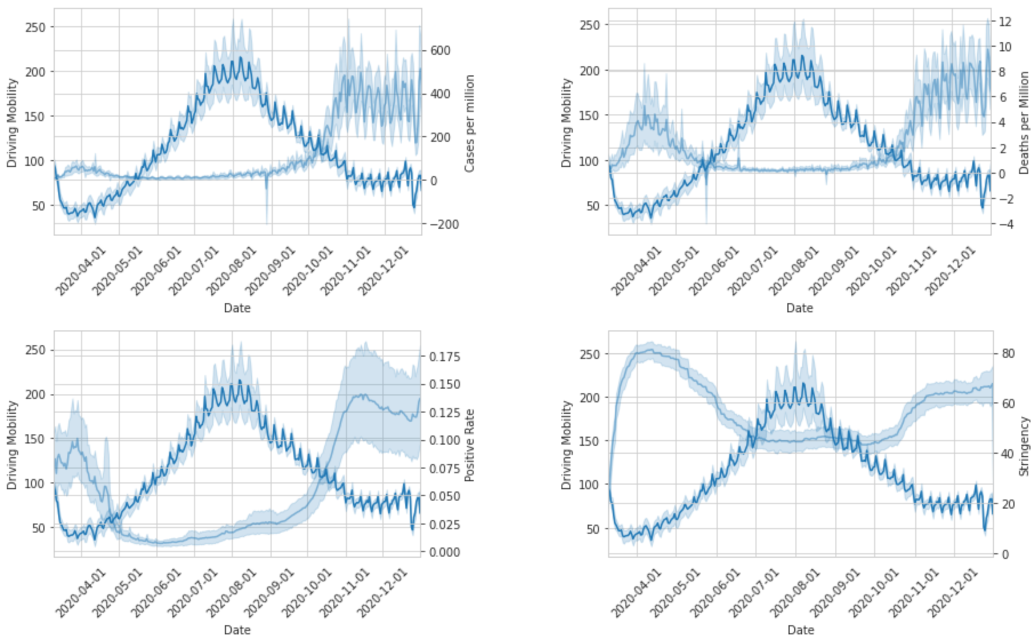


Figure 15: A comparison of driving mobility (Apple) with the number of cases (per million), deaths (per million), positive rate, and stringency index, average over the set of European countries.

18), i.e. the transit mobility also has a strong negative correlation with the stringency levels. Since the stringency levels do not change as frequently, we could use the mobility data to infer a country's stringency.

Next, we looked at comparisons of the Apple mobility data with the control response measures (from `2_ecdc/country_response_measures.csv`). The data in the file gave the start and end dates of the various response measures implemented across Europe. For example, this includes the closing of nurseries and banning outdoor events of more than 50 people. Since there were many different response measures and not all the same ones were implemented across countries, we looked at a more general perspective of how measures affect mobility. A selection of these plots is shown in the Appendix (Figure 20). A red line indicates that the country implemented a new restriction on that date, and a green dashed line is representative of the end of one of these measures. We include some observations of how mobility alters in relation to the level of response measures from these plots:

- Croatia: In August, a new restriction is introduced and driving mobility reduces. In November, a restriction is lifted, but mobility continues to reduce. This may indicate that behaviours were changed through nudges rather than specific restrictions.

- Sweden: As cases start to grow in September, mobility continues to reduce. When a restriction is lifted in October but cases are growing, people still reduce mobility. They did not require additional response measures to do so. This level of mobility is the same as the most strict regulations in March despite any additional measures being taken. When new measures are brought in in November, mobility does not reduce since it is already at its lowest level.
- Finland: In March, when Covid response measures were introduced, mobility reduced. Then, as the case number start to reduce in April, mobility begins to increase despite to lifting of restrictions. However, this does not increase the number of cases as expected. This suggests they had behavioural fatigue and were able to distance themselves appropriately without government intervention.
- France: There seems to be some non-compliance during the first lockdown. We also see reduced mobility as cases increase from August onwards, inspite of no new restrictions being invoked.

In an attempt, to removed this seasonal affect we looked at Google mobility data. This is due to the fact that the journeys can be sorted by purpose (e.g. grocery or recreational) and some of these purposes, like park visits, have an obvious seasonal trend and can be excluded from the analysis. We found that retail, transit, recreational and workplace were very closely correlated with $R(t)$ number generated from the SEIR model. Thus, we calculated the median value of journeys of these types to give us an average mobility statistic for each country. We also included the stringency index on these plots to give an indication of how strict the response measures are. Since we want to put 3 variables on the same graph, the stringency index is rescaled by a factor of 50 so that it can be plotted on the same axis as the reproduction number. These plots are shown in Figure 16.

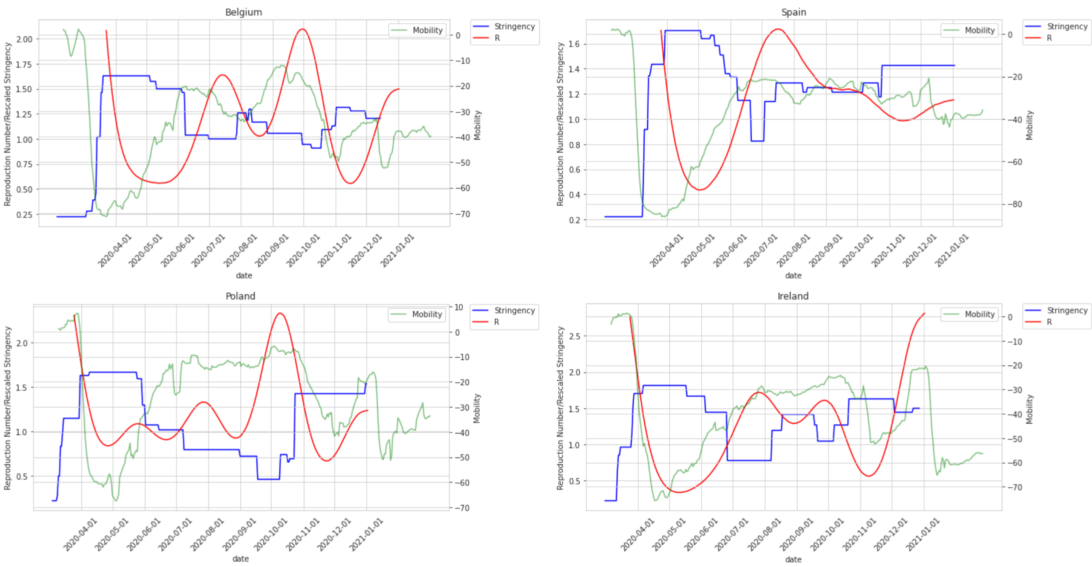


Figure 16: Google mobility data plotted with simulated $R(t)$ numbers and stringency index divided by a factor of 50.

For many countries, we see non-compliance with the restrictions. In Figure 16, for each of these countries, we see that the reproduction number increases after some time with strict levels of Covid response measures. For example, in Spain, during the highest stringency levels over Winter, the $R(t)$ number begin to grow after a while. This non-compliance could be due to the idea of behavioural fatigue. This can also be seen by the increasing mobility during the first lockdown, where stringency was very high. Non-compliance, again, increases the reproduction rate. Similarly, in the Poland data for the second lockdown that occurred in Winter, the reproduction number and mobility both begin to increase after a month of the high level of restrictions. Addressing the issues of non-compliance using ideas from behavioural science could be crucial in mitigating the effects of a pandemic.

If we now consider how mobility changed in Slovenia, we see that the highest Winter stringency was linked to mobility levels similar to the mobility seen in the first more stringent lockdown. This suggests that people reverted to their previous learned behaviours and only required a nudge rather than a strict policy to reach this. However, the same story is not true across all countries. In Spain, the lowest mobility levels seen in the first lockdown do not then occur later in the year. One possible explanation for this could be that Slovenia had lower restrictions during their summer period and, as such, did not experience as much behavioural fatigue as people living under stricter summer regulations in Spain.

2.8 Conclusions

We have made several contributions to the overall understanding of the development of the pandemic and its societal impacts.

Firstly, we recognised that the reproduction number $R(t)$ is of vital importance for understanding the spread of the pandemic. We found the estimates for the $R(t)$ values provided to us in the `owid/covid_data.csv` data set to be insufficient, as they are based on a simple SIR model, which does not take an incubation period of the virus into account. To remedy this, we proposed an SEIR model from which we derived our own estimates for $R(t)$ for each country.

We have used these $R(t)$ values to investigate non-compliance behaviour. We have found compelling evidence for progressive non-compliance in several countries as shown in figure 16. During the periods of lockdown, mobility and consequently $R(t)$ start to increase while restrictions are still in place, suggesting that people might not be following government imposed measures as closely as before. We invite the reader to investigate the phenomenon of progressive non-compliance in the accompanying R Shiny Covid predictor App under the"Cases and R0" tab.

We believe that predicting new cases of Covid as such is not the correct objective function when devising a strategy for combatting the pandemic. Rather than reducing the number of overall new cases, focus should be placed on reducing the number of severe Covid cases as measured by the number of Covid-related ICU patients. We how shown that ARIMA models leveraging external regressors can be a useful tool for predicting ICU cases for up to ten days in advance. We believe that such models can be a useful aid for forecasting demand on the national healthcare system. This may prove helpful to policymakers deciding whether new restrictive measures should be taken or not.

Notes

[1]Source: https://fullfact.org/health/boris-johnson-coronavirus-this-morning/

[2]Source: https://www.rijksoverheid.nl/documenten/toespraken/2020/03/16/tv-toespraak-van-minister-president-mark-rutte

References

Arroyo-Marioli, F., F. Bullano, S. Kucinskas, and C. Rondón-Moreno (2021). Tracking r of covid-19: A new real-time estimation using the kalman filter. *PLoS One 16*(1), e0244474.

Balmford, B., J. D. Annan, J. C. Hargreaves, M. Altoè, and I. J. Bateman (2020). Cross-country comparisons of covid-19: Policy, politics and the price of life. *Environ Resource Econ 76*.

Bürkner, P.-C. (2017). brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software, Articles 80*(1), 1–28.

Dowd, J. B., L. Andriano, D. M. Brazel, V. Rotondi, P. Block, X. Ding, Y. Liu, and M. C. Mills (2020). Demographic science aids in understanding the spread and fatality rates of covid-19. *Proceedings of the National Academy of Sciences 117*(18), 9696–9698.

Fancourt, D., F. Bu, H. W. Mak, and A. Steptoe (2020). Covid-19 social study. *Results release 15*.

Fraser, C., S. Riley, R. M. Anderson, and N. M. Ferguson (2004). Factors that make an infectious disease outbreak controllable. *Proceedings of the National Academy of Sciences 101*(16), 6146–6151.

Harvey, N. (2020). Behavioral fatigue: Real phenomenon, naïve construct, or policy contrivance? *Frontiers in Psychology 11*.

Hyndman, R. J. and Y. Khandakar (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software 26*(3), 1–22.

IEMAG, E. M. S. (2020). A population-level seir model for covid-19 scenarios. *https://assets.gov.ie/74595/e10ea7bb423c4110839e2f30c3817dc7.pdf*.

Marmot, M. (2005). Social determinants of health inequalities. *The Lancet 365*.

McAloon, C., Á. Collins, K. Hunt, A. Barber, A. W. Byrne, F. Butler, M. Casey, J. Griffin, E. Lane, D. McEvoy, P. Wall, M. Green, L. O'Grady, and S. J. More (2020). Incubation period of covid-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open 10*(8).

Michie, S., R. West, and N. Harvey (2020). The concept of "fatigue" in tackling covid-19. *bmj 371*.

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

Thaler, R. H. and C. R. Sunstein (2009). *Nudge: Improving decisions about health, wealth, and happiness.* Penguin.

Wang, H. and N. Yamamoto (2020). Using partial differential equation with google mobility data to model covid-19 in arizona. *arXiv preprint arXiv:2006.16928*.

Appendix

Dealing with negative case numbers

We noticed that several countries under investigation report a negative number of new cases on some days. We assume that this is to adjust for previous over-reporting. To be able to use these countries' data in our model we smoothed this negative number across the previous days in the following manner: Given a day t on which a negative case number was reported, we calculated the total number c_t of new cases in the previous two weeks, adjusted by that negative number. We then updated the previously negative number of new cases on day t to $1/14 \cdot c_t$. The remaining $13/14$ of c_t were spread out evenly across the remaining 13 days, keeping the proportions of reported days on each day the same.

This worked well for 27 out of the 28 countries under study. For Luxembourg, this approach did not work, since they had an overall low absolute number of cases and one adjustment that was large in magnitude, the result of which was that the negative number reported on that one day was larger than the sum of positive new cases reported in the previous two weeks. Therefore, we decided to spread out the negative adjustment over all the previous dates, keeping the proportions of reported cases on each date the same.

EDA: Mobility Figures

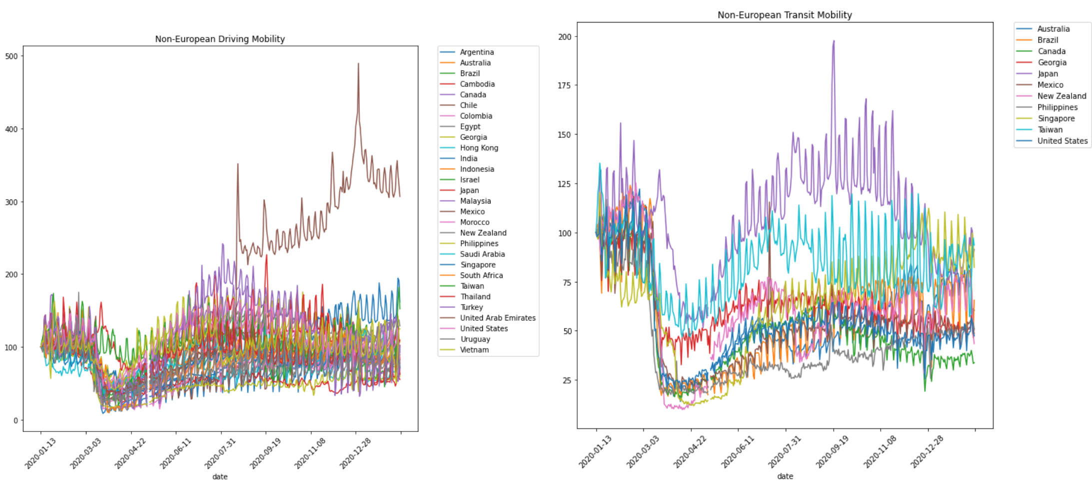


Figure 17: The Apple mobility data for non-European countries.

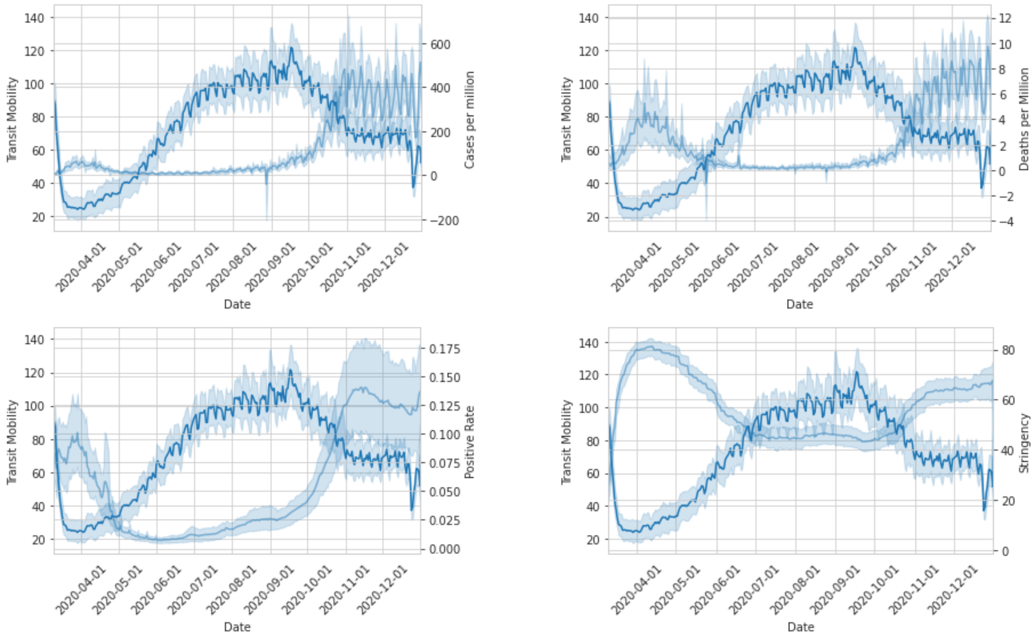


Figure 18: A comparison of transit mobility (Apple) with the number of cases (per million), deaths (per million), positive rate, and stringency index.

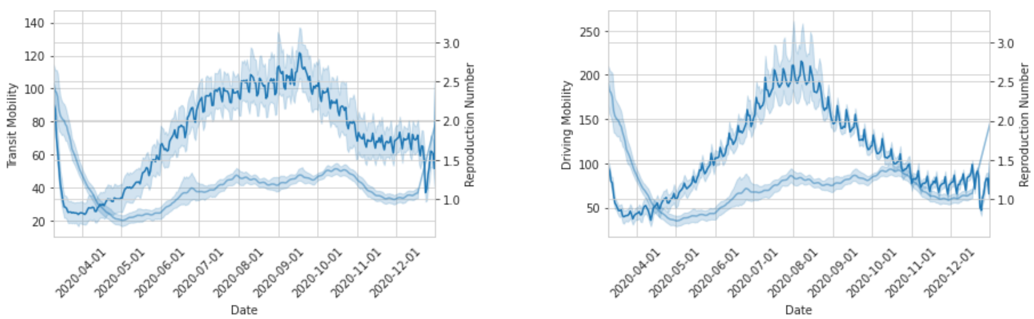


Figure 19: Transit and driving mobility (Apple) plotted with reproduction rate (Owid). There is a relationship between these two covariates. Thus, we show a comparison of our simulated $R(t)$ number of mobility data (Google) in the Shiny app.

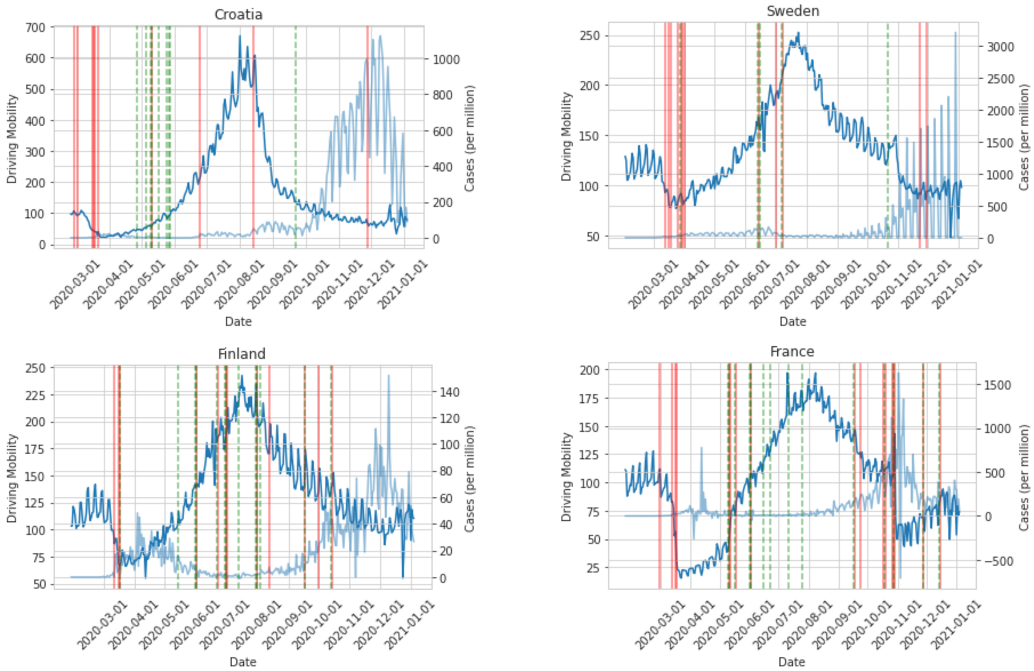


Figure 20: These plots show the driving mobility (Apple) and cases per million in relation to new country responses. The red lines indicate a new restriction has been introduced on that date and green dashed lines mean that a restriction has been lifted.