NTNU
Kunnskap for en bedre verden

Department Computer Science
TDT4259 Applied Data Science

Analysing the relationship between COVID-19 restriction policies and mortality

Authors:	*Student ID:*	*Email:*
Sara Johanne Asche	489768	saraja@stud.ntnu.no
Marie Burns Bergan	489767	mariebbe@stud.ntnu.no
Yuheng He	567187	hyhnaranja@gmail.com
Anna Katharina Herrmann	567012	annakher@stud.ntnu.no
Jenny Merkesvik	489783	jennm@stud.ntnu.no
Malin Stubban Berg	485500	malin.s.berg@stud.ntnu.no

November, 2021

Table of Contents

Abbreviations

BAM Business analytics methology. 5, 6

CIDSS Clinical Decision Support System. 6

COVID-19 Coronavirus Disease 2019. v, vi, 1–12, 15, 16, 18, 30–37

CRISP-DM The CRoss Industry Standard Process for Data Mining. v, 5–8, 11, 30, 36

DSR Design Science Research. 6

NPI Non-Pharmaceutical Interventions. 1, 8, 9

OxCGRT Oxford COVID-19 Government Response Tracker. 2, 3, 7, 12, 13, 35, 36, 42

PCP Pandemic Coronavirus preparedness. 9, 32, 35–37

PIP Pandemic influenza preparedness. 2, 3, 9, 32–34

SARS Severe acute respiratory syndrome. 10

SARS-CoV-2 Severe acute respiratory syndrome coronavirus 2. 1, 2, 5, 9, 10, 12, 15

UN United Nations. 1

WHO World Health Organisation. 1–3, 5–10, 32, 33, 37

List of Figures

List of Tables

1 Introduction

When the disease caused by the virus Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first discovered back in December 2019, little was known of the grave impacts it would have on global health and everyday life. After declared a pandemic in March 2020, the World Health Organisation (WHO) called for all countries to implement strict regulations to to limit the spread of SARS-CoV-2, and mitigate the adverse health outcomes of Coronavirus Disease 2019 (COVID-19). Examples include social distancing, public information campaigns, limiting travel within and between countries, closing workplaces and schools, and imposing quarantines and lockdowns. Despite these interventions, COVID-19 has claimed a total of 4.95 million lives worldwide as of October 2021 [1]. In addition, the regulations have had massive negative socioeconomic and health-related impacts. For instance, Mækelæ et al. highlight that the implementation of such policies and restrictions may have had detrimental effects on the well-being of individuals, with increased risk of developing anxiety, paranoia, and other mental disorders [2].

Many argue that the COVID-19 pandemic is ending, as a high percentage of the world population has been vaccinated. Norway has for instance had a rapid vaccination progress, with almost 86.6% of the adult population having received the second dose as of October 2021 [3]. However, there are currently many countries that are still administrating the first doses to their inhabitants. In addition to the mortality caused by slower vaccinations, the continued spread of the virus also enables new SARS-CoV-2 variants to emerge. These new variants may omit the immunity given by the current vaccines, potentially causing another global COVID-19 outbreak. In November 2021, there has been an increase in the number of cases reported, despite vaccines being distributed all over the world: between November 8^{th} and 14^{th}, 3.3 million cases were reported, representing a 6% increase from the previous week [1]. These numbers indicate that the pandemic is not coming to an end yet.

Furthermore, many re-emerging diseases other than COVID-19 also exist, and new diseases still emerge and evolve. Frieden et al. highlights the importance of being prepared for the next pandemic, which is inevitably coming [4]. With any new pandemic, developing pharmaceuticals and vaccines to hinder mortality takes time. It is therefore of importance to study Non-Pharmaceutical Interventions (NPI), which may be implemented as soon as a new viral threat is identified. With these points in mind, it becomes apparent that analysing and understanding epidemiological data from the past two years can provide valuable insight.

1.1 The World Health Organisation and the Pandemic influenza preparedness and response

Throughout the COVID-19 pandemic, the WHO has been an essential advisor on the implementation of policies and restrictions, both nationally and globally. The WHO is a United Nations (UN) agency founded in 1948, responsible for monitoring international public health. Its mission is to work "*worldwide to promote health, keep the world safe, and serve the vulnerable*" [5]. To improve universal health coverage, WHO strives to increase access to quality primary health care, essential medicines, and other health products and services. The work of WHO is defined in its constitution, and is divided into three different categories [6]:

1. Normative functions
2. Directing and coordinating functions
3. Research and technical cooperation functions

The *normative functions* (1) include the administration of international conventions, agreements, and regulations; non-binding standards; and recommendations. Meanwhile, the *directing and coordinating functions* (2) are broader functions describing the WHO's work towards health for all, poverty mitigation, and development of disease- specific programs. Finally, *research and technical cooperation functions* (3) consider disease eradication and the handling of global health emergencies [6]. In global health emergencies like the COVID-19 pandemic, the tasks of WHO include detecting

and responding to emergencies; supporting the development of necessary tools; and delivery of essential health services in fragile settings. WHO has also served as a global hub of information regarding COVID-19 cases, vaccination progress, and mortality rates. The information has been collected from various national health institutions and has displayed on its websites throughout the pandemic.

In response to the Swine Flu in 2009, the WHO published the framework *Pandemic influenza preparedness (PIP) and response* [7]. This document is an international instrument that details how to deal with different phases of the reoccurring influenza pandemic. Table 1 presents the nine periods that simulate how a pandemic usually unfolds in real-life scenarios: phases 1-6, post-peak period, possible new wave, and the post-pandemic period.

Table 1: Overview of the phases of an influenza pandemic, according to the WHO framework *Pandemic influenza preparedness and response*.

Phase	Description
1	No novel influenza virus circulating among animals have been reported to cause infection in humans.
2	A novel Influenza virus circulating in domesticated or wild animals is known to have caused infection in humans.
3	A novel influenza reassortant virus has caused small clusters of disease in people, but not resulted in human-to-human transmission sufficient to sustain community-level outbreaks.
4	Human-to-human transmission of an novel influenza reassortant virus able to sustain community-level outbreaks has been verified.
5	The same virus has caused sustained community-level outbreaks in two or more countries in one WHO region.
6	In addition to the criteria defined in Phase 5, the same virus has caused sustained community-level outbreaks in at least one other country in another WHO region.
Post-peak period	Levels of pandemic influenza in most countries with adequate surveillance have dropped below peak levels.
Possible new wave	Level of pandemic influenza activity in most countries with adequate surveillance rising again.
Post-pandemic period	Levels of influenza activity have returned to the levels seen for seasonal influenza in most countries with adequate surveillance.

The PIP framework utilises the nine pandemic phases in order to function as a planning tool, with tasks and recommendations for each phase. A similar framework is yet to be described in detail for the current COVID-19 pandemic, which is predicted to be re-emerging just like the influenza virus [8]. Therefore, it would be in the interest of WHO to look into what this phase plan would look like for viruses of the Coronaviridae family.

1.2 The Oxford COVID-19 Government Response Tracker: A measurement of COVID-19 related policies and regulations

The rapid spread of SARS-CoV-2 prompted governments worldwide to implement various policies. These aimed to mitigate of transmission and death; aiding individuals, businesses, and other entities in need; and establishing prioritisation plans for testing and vaccination, among others. Tracking and classifying these measurements is the main aim of the Oxford COVID-19 Government Response Tracker (OxCGRT), a project led by the University of Oxford [9]. Throughout the project, policies implemented in 186 countries have been tracked since January 1st 2020. They are separated into different categories depending on their intended effect, and the data is captured with the help of volunteers specifically trained to ensure consistent data. This has produced a large dataset on how policies and regulations have varied in each country in the close to two-year duration of the COVID-19 pandemic. The tracker offers valuable insights into government responses, and allows for the analysis of their importance for the occurrence of COVID-19 infections and deaths.

1.3 Motivation and problem definition

The initial interest for analysing data related to COVID-19 originates from how it has affected our everyday lives for the past two years; both in terms of the looming disease itself, as well as the policies and restrictions that have been implemented. In addition, four of the group members are Biotechnology students familiar with the field of virology. Sara and Marie are especially invested in the topic as both their master's theses regard the prediction of COVID-19 spread. Based on the team member's backgrounds and the timeliness, we decided to collect and analyse COVID-19 data.

In light of the OxCGRT dataset, it would be interesting to investigate government strictness and its impact on the COVID-19 pandemic. One possibility is to investigate whether or not governmental strictness is related to the mortality rates of COVID-19, in an attempt to see if the restrictions have helped in mitigating mortality. Furthermore, the restrictions with the highest impact on mortality could be identified. With this information, the WHO could recommend governments on how to respond to COVID-19 outbreaks without implementing more or stricter policies than needed. Because strict regulations have had such grave impact on the health, well-being, and economy across multiple countries, companies, and individuals, the WHO could also advise against some restrictions if they are found to be less effective. Furthermore, knowing the effectiveness of regulations and restrictions could help categorise them, so they may be applied to a Coronaviridae version of the PIP framework. Thus the data collected on restriction levels and mortality could be used to answer the defined problem: **Which restrictions and policies have been the most effective in mitigating COVID-19 mortality?**

The stakeholders for the project's defined problem are many, because COVID-19 restrictions have been relatively invasive and costly overall, as discussed in Section 1.1 (p. 1). Therefore, many entities are affected by the chosen pandemic approach. There are four main stakeholders. Firstly, the governments of countries interested in implementing COVID-19 restrictions. It is in their interest to reduce the costs associated with restrictions, and as well ensure the health and well-being of the country's inhabitants. Second, businesses affected by the implemented restrictions. Examples include the closing of workplaces, customer losses due to stay-at-home requirements, and economic support policies. The third stakeholder is the general population, which is affected by the invasiveness and loss of autonomy that may be associated with some restrictions. The final stakeholder is the WHO, which is an essential advisor during health emergencies. Having up-to-date information on which restrictions are effective in mortality mitigation is therefore in the interest of WHO. Furthermore, the project can also provide useful information to the WHO on which policies to continue recommending, and which can be exclude from for instance a Coronaviridae version of the PIP framework. The WHO's aim of "health for all" can also be achieved by avoiding the detrimental effects of unnecessarily invasive restrictions on mental health.

1.4 The team

The team behind the present work is a multidisciplinary group of six students from three different study programs at NTNU. They share some common knowledge, but also possess individual expertise that can benefit the team and their project. A short overview of the background and competence of each group member is provided in the following subsections.

1.4.1 Anna - M.Sc in Applied and Engineering Physics

As part of her master's program in Applied and Engineering Physics at the Technical University of Munich, Anna is currently doing an ERASMUS semester at NTNU. During her bachelor's degree in physics at the KIT, she has gained experience in applying machine learning to research problems and analysing large amounts of data using statistical methods.

1.4.2 Jenny - M.Sc in Systems Biology

As a student of Biotechnology, the field of Data Science is relatively new to Jenny. The specialisation in Systems biology has however introduced her to Bioinformatics and thus provided experience in the handling of large amounts of data. This skill is developed further in Jenny's current Master thesis, in which she is collecting microbial traits and genomic sequences, with the aim of utilising machine learning in genotype–phenotype association for hitherto uncultivated microbes.

1.4.3 Malin - M.Sc in Systems Biology

Malin studies Biotechnology in her fifth year, also with a specialisation in Systems Biology. She is currently working on a master thesis focusing on the integration of microbiota using supervised machine learning. The area of Data Science is quite new to her, however, she has some background in the field of Bioinformatics and how to handle large data sets.

1.4.4 Marie - M.Sc in Systems Biology

Marie is a fifth year student at the Biotechnology program with a specialisation in Systems Biology. She is not very familiar with Data Science, but has however developed an interest for modelling and epidemiology through her master thesis. Currently, she is working on predicting the spread of COVID-19 in health care facilities with an agent based model.

1.4.5 Sara - M.Sc in Systems Biology

Sara is on her final year of a M.Sc in Biotechnology, with Systems Biology as her chosen specialisation. She is currently writing her thesis about predicting the spread of COVID-19 using an agent based model, and is therefore well traversed in the topic of viral disease and working with epidemiological data.

1.4.6 Yuheng - M.Sc in Software Engineering

Yuheng is on his second year of a master in Software Engineering at the University of Seville. He has previous experience with machine learning and Cloud. Currently, he is doing his thesis on the application of machine learning algorithms in data mining.

1.4.7 Team member roles and contributions

An overview of the roles and contributions of the members is given in Table 2. Additionally, all have participated in exploring and deciding the project topic, data sources, problems, and objectives.

Table 2: Roles and contributions of all team members.

Anna: statistics and analysis	Jenny: datasets and interpretations	Malin: data strategy
Python code for mortality shift. Investigate effects of mortality shift. Analysis section. Statistics consultant.	Dataset preparation and assembly. Methods section on datasets. Interpretation section. Recommendations and implementation section. Visualisations with figures and tables. Presentation manuscript, slides, and recording.	Introduction section. Background section on CRISP-DM. Limitations and future work section.
Marie: virology	**Sara: virology and analysis**	**Yuheng: statistics and analysis**
Introduction section. Background section. Limitations and future work section.	Introduction and Background sections. R code for Wilcoxon and Spearman analyses. Methods and Analysis sections on R. Reccomendations and implementation section. GitHub setup. Presentation manuscript and slides.	Exploratory analyses in Python. Statistical background on Shapiro-Wilk test.

2 Background

2.1 Early project-defining choices

Before a detailed project plan was established, the team conducted several brainstorming sessions to initiate the project work. This subsection details two early decisions made: those of the main problem the project sought to address, and which strategy should be utilised to achieve it.

2.1.1 Choice of project problem

There is a wide range of available data related to the COVID-19 pandemic, and several possible topics were considered for the project. First, an investigation of the evolution of different SARS-CoV-2 variants was discussed. Contemplated topics were how the variants spread, and whether the relative abundance of variants could be predicted given an initial variant distribution. Due to the lack data sufficiently covering variant abundances, the idea was discarded. Secondly, the team considered analysing the effectiveness of the vaccination strategies in different countries. However, the value of conducting such a study to benefit the WHO would be limited, seeing as the countries with the most vaccination data available usually had the same vaccination strategy. Therefore, a third option was proposed: it regarded the use of machine learning to predict mortality rates based on restriction stringency level in selected countries. This idea was however discarded. Mortality and strictness prediction would not be very valuable this far into the pandemic: the fraction of vaccinated individuals is increasing steadily, whilst restrictions and policies are being lifted.

Instead, the team decided to conduct an extensive analysis of the relationship between stringency and mortality, defined as COVID-19 related deaths per number of confirmed COVID-19 cases. However, looking into all countries that had available data on mortality and stringency levels would not be a good way to approach this problem. The data from the beginning and towards the end of the pandemic contains much variation for each country, which impedes the ability of drawing any conclusions from a simple mean of all countries. Thus the team decided to categorise countries with similar mortality and stringency trends, with the aim of using these to reach more specified conclusions. Thus the primary goal was reduced to use the analysis of selected data to investigate whether certain restrictions can be altered or removed without affecting the mortality rate. This knowledge could be helpful for the WHO to provide a framework on how to approach the different phases of pandemics caused by viruses of the Coronaviridae family.

2.1.2 Choice of data-driven strategy

The The CRoss Industry Standard Process for Data Mining (CRISP-DM) strategy was chosen as the methodology for the project. The main reasoning behind this choice of strategy above other alternatives was that CRISP-DM has been well documented, and was considered as well suited for the defined problem.

The first version of CRISP-DM was presented and published in 1999 [10]. Since then, CRISP-DM has become the most widely used framework for data mining. It is a cost-effective methodology due to the number of phases that must be overcome before commencing more costly project phases, and due to its encouragement of best practice and replicability. This is further emphasised by the strategy's flexibility. Models are improved through each iteration of the framework, thus making it easier to initiate a project despite lack of prior domain knowledge. Furthermore, the cross-industry aspect of CRISP-DM makes is applicable to any data science project, regardless of project domain. This includes data science projects like the present work, where the main objective is not to increase capital with focus on customers, but rather to perform a diagnostic analysis.

On the other hand, the Business analytics methology (BAM) framework seeks to align a business analytics project with a business goal throughout each step of the analysis. Hindle and Vidgen describe that one of the strengths of BAM is its ability to connect *"analytics with an organisation's thinking regarding purpose, strategy and core activities*", which is used to *"help an organisation to*

create business value" [11]. BAM was found to be appropriate for the third sector in particular, where for instance the WHO would be placed. However, the team decided against using the BAM as it focuses heavily on creating business value [11], in addition to the lack of flexibility throughout the framework's four stages. The WHO is already funded by governments, and therefore not in need of customers nor donations to increase its capital or business value. Furthermore, the team's aim is to provide recommendations to the WHO regarding the third category in its constitution, *research and technical cooperation functions* (see Section 1.1, p. 1). Therefore, many of the approaches used in the BAM framework, like the business model canvas (BMC); activity model; and customer, actor, transformation, worldview, owner, and environment (CATWOE) analysis, would not be of interest for this project. However, the team decided to implement BAM's analytics leverage matrix for its recommendations to the WHO, as it provided a clear way of displaying and interpreting the results of the conducted analysis. The matrix can be found in Section 5.2 (p. 32)

Data-driven design thinking is another example of a possible project strategy. It largely focuses on understanding users or costumers in order to create a valuable product or prototype. Such an approach would be difficult to implement in the present work, as the goal is to provide advise for the WHO which are forwarded to governments all over the world. This entails that the current project is more of a diagnostic investigation on how COVID-19 mortality is influenced by the introduction of various policies, rather than the production of a tangible or implementable model for the WHO.

Thus the CRISP-DM was believed to be a strategy that complimented the project aims, as well as providing the necessary flexibility in business understanding. Still, the utilisation of a leverage matrix highlights the underlying similarity between the present work and newer business methodology approaches. Ferreira et al. [12] and their Clinical Decision Support System (CIDSS) is an example of another data-driven project that has utilised CRISP-DM within the domain of COVID-19. The CIDSS project used CRISP-DM alongside Design Science Research (DSR) to create predictions regarding the most likely outcome for a COVID-19-infected individual, whether that be death or recovery. The predictions could further be utilised to assist clinicians in crucial decision-making moments where resources are extremely scarce [12].

2.2 The CRISP-DM framework

This section provides a general introduction to the CRISP-DM framework, which is the chosen methodology for the present work. This structured approach is used for planning data mining projects, and is broken down into six phases. Even though the following subsections describe the phases in a linear manner, an important note is that the phases are not rigid: moving back and forth between the steps is required when following the method. In addition, the model has a cyclical nature, where each phase is improved or developed further through multiple iterations. As will be evident throughout the report, the methodology of the present project indeed shows the mobility back and forth between steps, as well as iterations, by for instance conducting one analysis/model, considering the results, and including more data in another analysis.

2.2.1 Business understanding

The initial phase of CRISP-DM is *business understanding*. It focuses on identifying project objectives from the business' point of view. Requirements from and knowledge on the business must be taken into account to achieve this. Chapman et al. highlight how the business objectives can be defined by what a customer wants to accomplish [10]. The primary objective of the customer should then be considered from a business perspective. Furthermore, as these objectives are given in business terminology, they must be formulated into a data mining problem in technical terms. The data mining problem must contain the data in which the objective is based on, as well as the possible outcome of the problem. An example is given by Chapman et al., in which the data mining goal is as specific as *"Predict how many widgets a customer will buy, given their purchases over the past three years, demographic information (age, salary, city, etc.), and the price of the item*" [10].

Furthermore, a project plan should be outlined and described within this first CRISP-DM pase. It describes the steps that ought to be performed throughout the project, thus providing a time schedule for when iterations of CRISP-DM should occur, as well as when different project milestones should be reached [13]. Such a plan can thus help in the understanding of the broad project context, in addition to narrowing its scope.

In the present work, the *business understanding* consisted of gaining domain knowledge of the project topic, which included COVID-19, the WHO, and COVID-19 regulations and policies (see Sections 1.1-1.2, pp. 1-2). A preliminary plan was designed based on this knowledge and the problem formulation (see Section 1.3, p. 3), and is further detailed in Section 2.3 (p. 8). This section also includes the formulation of the project objectives. From these, their data mining formulation were obtained by considering the possible data sources available. An example is the reformulation of *"identifying a possible correlation between [COVID-19 policy] stringency and COVID-19 mortality*" to *"identifying a possible correlation between [COVID-19 policy] stringency and COVID-19 mortality, utilising the OxCGRT dataset fields on confirmed cases, deaths, and stringency*".

2.2.2 Data understanding

The second phase of CRISP-DM is *data understanding*, which entails collecting, describing, and exploring of data, as well as verifying its quality. Examples of properties to assess are accuracy in terms of what it describes; relevancy to the project; completeness, in terms of few missing entries; timeliness to the intended application; and consistency, to ensure proper data format [14]. As well during this phase, a hypothesis can be formed based on interesting findings during the data exploration [13]. Because the CRISP-DM framework is not rigid, these observations can then be used to further refine or specify the problem and its objectives.

For the project regarded in this report, different datasets were considered before deciding which to proceed with. The process of data understanding is closely related to *business understanding*, as outlining the data-mining problem and the project plan required understanding of the available data [13]. Thus in this phase, the team made an effort to ensure that all members understood and agreed on the choices by going through the datasets. The process of identifying and assessing data is described in Section 3.1 (p. 11).

2.2.3 Data preparation

The third CRISP-DM phase, *data preparation*, covers the process of assembling relevant data into one dataset. Firstly, the data needs to be cleaned, a process often involving deleting or inserting defaults for unrepresentative data. New attributes may be constructed based on the available fields, and the data may be transformed ahead of the modelling phase, if necessary. If multiple datasets are used, the data needs to be integrated into a merged dataset [13]. The process of dataset cleaning and assembly is detailed in Section 3.1 (p. 11).

2.2.4 Modelling

The fourth phase the CRISP-DM approach involves selecting appropriate modelling techniques for the defined objectives, as well as selecting relevant parameters. There are often several techniques that can be applied for the same data mining problem. This step sees the choosing of one appropriate for the chosen data. After selecting a modelling technique, the next steps are to generate a test design; build a model; and lastly assess the model. During this process, problems with the data may be discovered. Thus this phase is closely linked to the *data preparation* phase [13]. The analyses and models used to address the defined problem are presented in Section 3.2 (p. 3.2).

2.2.5 Evaluation

In the *evaluation* phase, the results of the models are evaluated before the entire process is reviewed, and the next course of action can determined. During the evaluation of the results, how the model meets the business objectives should be assessed. If the conclusion of this assessment is that the model is unsatisfactory, the process should be reviewed in order to identify whether any of the previous steps were overlooked or insufficiently implemented. Further, the model as a whole should be reviewed once over to assure its quality. Finally, when the model is deemed appropriate, the next phase can be commenced. [13]. For the current project, the model results are interpreted in Section 5.1 (p. 30), while its limitations are regarded in Section 5.3 (p. 35).

2.2.6 Deployment

The last phase of the CRISP-DM framework is *deployment* of the findings. It involves plans for the deployment, monitoring, and maintenance. A final review of the project should also be conducted. During this phase, technical terminology should be converted back into business terminology, which is more appropriate when the stakeholders are presented with the results and how they can benefit from the project. The steps of *deployment* are in many cases carried out by the users, not the data analysts. Regardless, it is important for the data analysts to be aware of the actions needed to make use of their model, in addition to the need for monitoring and maintenance of the deployed plan [13]. The recommendations and implementation plan resulting from this project are found in Section 5.2 (p. 32).

2.3 Objectives and Project plan

As mentioned in Section 1.3 (p. 3), there are several stakeholders affected by the challenge of finding appropriate stringency levels, in terms of policy types and numbers, to decrease mortality without compromising welfare. In order to answer the defined problem of finding which restrictions and policies have been the most effective in mitigating COVID-19 mortality, three objectives have been formulated:

1. Is there a correlation between stringency and COVID-19 mortality?
2. Which policies have contributed to reducing COVID-19 mortality?
3. How did the vaccine impact COVID-19 mortality compared to non-pharmaceutical interventions?

The objectives seek to divide the problem investigation into more tangible steps. The first objective regards the fact that implementing policies and regulations have been the main strategy for COVID-19 mortality mitigation. Establishing whether their strictness is related to COVID-19 mortality is important to evaluate the approach of governments and the WHO during the current pandemic. Furthermore, it is necessary to establish the existence of a relationship between stringency and mortality for the following analyses.

The second objective seeks to assess the relative effect of the various policies on mortality. It thus provides insight into which types of policies have been effective, and which can be neglected due to a lacking negative effect on mortality. Additionally, it may indicate whether a country's chosen strategy for handling the pandemic has been successful or not. For instance, a situation where the same policy has shown varying success in two countries could indicate which approach have been better.

Lastly, the third objective focuses on an investigation of whether NPIs can achieve as significant reductions in mortality rates as vaccinations. This is especially important for finding the best way of mitigating mortality in the early stages of a pandemic, before appropriate pharmaceuticals may have been developed.

Together, these objectives should provide results of relevance to all defined stakeholders. Individuals, businesses, governments and the WHO alike can benefit from knowing whether restrictions help mitigate COVID-19 mortality; which restrictions in that case are the most effective (and thus also which can be disregarded in some phases of a pandemic); and finally the relative effectiveness of the NPIs and vaccinations. On a broader scale, achieving the objectives could aid the WHO in preparing a "Pandemic Coronavirus preparedness (PCP)" framework, much alike the PIP framework for Influenza viruses. The business success criteria of the project is therefore to be able to formulate this framework based on the results of the analysis.

To answer the defined problem, the team decided on the following simplified version of a project plan:

1. Gain domain knowledge on the WHO and the COVID-19 pandemic.
2. Acquire relevant data to perform the investigation of the relationship between COVID-19 mortality and implemented policies.
3. Define groups of countries with similar demographics and/or disease history from this or previous pandemics.
4. Select one representative country from each group.
5. Compare and interpret how different restrictions affects COVID-19 mortality in the representative countries.
6. Formulate recommendations to the WHO as to which restrictions to implement first, in the form of a framework similar to the PIP framework.

The team has a time span of eight weeks to conduct the project, from September 22$^{\text{th}}$ until November 22$^{\text{th}}$, 2021. A general outline for the project plan's six steps is loosely defined. Steps 1-3 are covered in the Introduction (Section 1, p. 1) and Background (Section 2, p. 5) sections, and should be completed within the first four weeks of the project. Step 4 should be decided during the fifth project week. Steps 5 and 6 are reserved for the final three weeks of the project.

2.4 Grouping countries for analysis

During the current COVID-19 pandemic, countries have had different strategies in selecting NPIs, both in terms of *which, when,* and *how*. This makes analysing epidemiological data surrounding restrictions and policies on a global scale challenging, as both the gravity and type of restrictions varies between countries. Instead of identifying trends in restrictions related to the global mortality rate, analyses and conclusions are based on groups of countries that have similar demographics and/or epidemiological history. Based on these criteria, the groups can be analysed and used to draw conclusions as to how different restrictions and policies may affect mortality.

2.4.1 Heavily impacted European countries

After the discovery of SARS-CoV-2 in December 2019 [15] in Wuhan, China, it did not take long for the virus to spread across national borders. In the early stages of the pandemic, some European countries were suffering from a large first wave of COVID-19 cases, hospitalisations, and deaths. These countries include Italy, France, Spain, and the United Kingdom (UK). In addition to their large initial waves, the countries share some demographic characteristics. Within this group, Spain and UK have the smallest and largest populations of 47 million and 67 million, respectively. The mean age within the group ranges from 47 in Italy, to 40 in UK. Finally, the urban population is around 80% for France, Spain, and UK. Meanwhile, Italy has the lowest percentage of urban population with 69% [16].

One reason for this experience may be that they had a relatively widespread circulation of the virus before restrictions were implemented. However, multiple factors may also have been at play. For instance, Italy has the most elderly population in Europe [17]. Still, analysing the impact of policies on mortality in these countries with a high initial spread of COVID-19 could be used to assess the need for restrictions in earlier phases of a possible pandemic.

2.4.2 Nordic countries

The Nordic countries Norway, Finland, and Denmark shared similar approaches in limiting transmission of SARS-CoV-2 nationally. They also have similar demographic characteristics. All countries have a population around 5.5 million, with mean ages ranging from 40 in Norway to 43 in Finland. In addition, the countries have a similar percentage of urban population, only ranging from 83% in Norway to 88% in Denmark. Therefore, it is reasonable to group the Nordic countries together [16]. Sweden was excluded as their strategy throughout the COVID-19 pandemic has been quite different to that of the other Nordic countries.

Most group members were situated in Norway during the pandemic, and all have experienced the Norwegian government's handling of restrictions. The local news are also biased towards covering the events and trends in other Nordic countries. This provides valuable insight into the situations of these countries in terms of restriction policies and mortality. It also sparks some curiosity in terms of inferring the effectiveness of their similar approach.

2.4.3 Countries affected by the 2002 SARS outbreak

Back in November 2002, a novel coronavirus was recognised in the Guangdong Province, China. It was named Severe acute respiratory syndrome (SARS) due to the pneumonia-like disease associated with it. The virus spread to Hong Kong and 30 other countries within February 2003. In coordination with WHO, a large-scale mobilisation to mitigate the spread of the virus was initiated. As a result, SARS was contained in less than four months after its discovery. However, the mortality of SARS was almost 10%, and East Asian countries were amongst those hit the hardest.

Due to their encounter with SARS in 2002-2003, it is possible that many East Asian countries responded much stricter during the early phases of COVID-19. Thus it would be of interest to look into the relationship between the policies and mortality in these countries, and how it potentially differs from countries without previous outbreaks of a highly contagious respiratory disease. Hong Kong, Malaysia, South Korea, and Thailand were chosen as representatives for this East Asian group. This grouping does not share as similar demographics as the previous groupings, due to their categorisation being based on their epidemiological history. The countries' populations vary from 7.4 million in Hong Kong, to 69 million in Thailand. The countries' mean ages are also quite different: Malaysia has the lowest mean age of 30, while the highest mean age of 45 is found in Hong Kong. There is also large differences between the countries percent of urban population, ranging from 51% to 82% in Thailand and South Korea, respectively. These observations may result in more variation between these countries than that observed for other groups.

2.4.4 Central Europe: low and high mortality groups

There have been some differences in how countries in Central Europe have approached the current pandemic. In the process of reviewing these countries, it was discovered that a few countries reported much higher mortality rates than others. Thus, two groups have been created for Central Europe. The lower mortality group includes Austria, Czechia, Germany, and Poland. Meanwhile, the higher mortality group comprises of Belgium, Hungary, and the Netherlands.

When it comes to the demographics of the high mortality countries, the mean age is 43. However, their populations range from 9.6 million in Belgium to 17 million in the Netherlands. Finally, Belgium and the Netherlands both have a high percentage of urban population above 90%, whilst Hungary only records 72%. Despite these differences, the countries are grouped together based on their geographical position in Central Europe, as well as their relatively high mortality compared to the other Central European group.

Similarly to the high mortality countries, the low mortality countries have an mean age of around 43; and a large variation in population size, from 9 million in Austria to 83.7 million in Germany. Austria and Poland have around 60% urban populations, whilst the percentage for Czechia and Germany is 75%. Overall, comparing these two groups of somewhat similar Central European

countries could make it possible to extract information as to why one group has a higher mortality than the other.

3 Method

As described in Section 2.1.1 (p. 5), CRISP-DM was chosen as the data-driven strategy for the project. While the first step on *business understanding* and objective identification (Section 1, p. 1; and Section 2.3, p. 8, respectively) has been described previously, this section will detail the approach for the next three steps. Figure 1 describes the overall workflow.

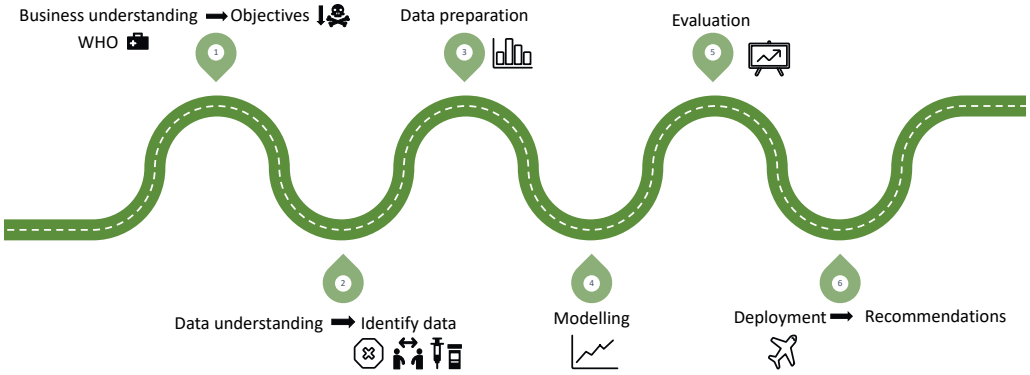

Figure 1: The six steps of the CRISP-DM strategy, with additional figures and keywords representing the process of the current project.

For the second step of the approach, the datasets presented in Sections 3.1.1-3.1.5 (pp. 11-14) were identified as helpful to answer the project objectives. Further, the step of *data preparation* was conducted as described in Section 3.1.6 (p. 14). The reasoning behind the chosen modelling approach is presented in Section 3.2 (p. 14). Meanwhile, the Evaluation and Deployment plan are reserved for Section 5.

3.1 Datasets

The analysis of the present work is based on a dataset that combines fields from five different sources. All raw data sheets can be found in the project's GitHub (see Appendix C, p. 45), while their overviews are described in Appendix A (p. 40). The following subsections include descriptions of the raw data sources and fields found to be relevant for the project. The data quality was further evaluated based on accuracy, relevance, completeness, timeliness, and consistency [14]. Finally, the selected data fields were assembled into the master dataset *combinedCovidData*, completing the third step of CRISP-DM.

3.1.1 COVID-19 World Testing Progress

The dataset *COVID-19 World Testing Progress* [18] originates from Our World in Data, and contains several fields of data on testing, registered per country per day, for a total of 138 countries. All fields in the raw dataset can be seen in Appendix A.1 (p. 40).

The accuracy and formatting consistency of the dataset has been evaluated as good based on the source and look of the dataset, respectively. The completeness varies between countries, although this depends on when each country experienced their first COVID-19 cases, as well as the availability of tests. These factors also influence the timeliness of the data. It is generally updated daily and ranges from January 1st 2020 to September 28th 2021, the latter of which was when the dataset was downloaded. Overall, the dataset was deemed of appropriate quality for the intended use.

The field included in *combinedCovidData* is the number of daily tests per thousand inhabitants, in addition to the date and country to allow integration with the other datasets.

3.1.2 COVID-19 Variants Worldwide Evolution

The dataset *COVID-19 Variants Worldwide Evolution* [19] contains entries per country per date per variant, detailing the relative abundance of 21 named SARS-CoV-2 virus variants. See Appendix A.2 (p. 40) for an overview of the full dataset.

The data originates from Our World in Data, thus it is believed to be accurate. In terms of timeliness, entries range from May 11$^{\text{th}}$ 2020 to September 16$^{\text{th}}$ 2021, meaning the week before the dataset was downloaded. New entries are added about every two weeks, thus this is a smaller dataset than those updated daily. This is also true for its coverage, which is 95 countries. Still, the data is consistent and relevant, and thus deemed appropriate for its intended use. However, the format of raw dataset was changed, as its entries were recorded per country per date *per variant*. By partially transposing the dataset, the variants were given as columns rather than as rows. This yielded entries per country per date, matching the other datasets utilised in *combinedCovidData*.

3.1.3 COVID-19 World Vaccination Progress

The dataset *COVID-19 World Vaccination Progress* [20] includes a data sheet called *Country vaccinations*. It details the vaccination progress per country per day. See Appendix A.3 (p. 41) for a description of the full data sheet.

The content of this sheet also originates from Our World in Data, which is believed to provide accurate entries. It covers 219 countries and most have daily data entries. However, not all fields are included in these daily updates for all countries. In terms of timeliness, the entries range from December 2$^{\text{nd}}$ 2020 until August 8$^{\text{th}}$ 2021, thus this dataset ends just over a month before the other datasets. Despite these observations, the completeness and coverage are both appropriate for the intended use of the data. The entries are consistent in format, and can be assembled into *combinedCovidData* using the *country* and *date* fields as keys. Of relevance for the project are the the number of daily vaccinations and fully vaccinated individuals, both per thousand inhabitants.

3.1.4 Oxford COVID-19 Government Response Tracker

The dataset *Oxford COVID-19 Government Response Tracker* [9] is a dataset detailing COVID-19 restrictions, policies, and various economic investments per country per day, for 186 countries. A full overview of the fields and their possible values is included in Appendix A.4 (p. 42).

As described in Section 1.2 (p. 2), this dataset has been created by trained individuals gathering information from various agencies and sources native to each country. Many of the contributors are specialists in the country they report on, which allows the consideration of news and sources in native languages. Only the policies are recorded however, thus no information on the concrete implementations can be gained from the dataset. The accuracy and consistency of the data are believed to be good. Further, the timeliness and completeness are both of high standard, with little to no missing data for the entries that range from January 1$^{\text{st}}$ 2020 until the download date, September 28$^{\text{th}}$ 2021.

The set contains 24 different indicators, split into three categories. 20 of these were included in *combinedCovidData*. The indicators are utilised in the calculation of a total of five indices, four of which are included in the present work: stringency, government response, health containment, and economic support. They provide a general overview of the extent of the government activity, per day per country. Each index is determined by a specific subset of relevant indicators. Lastly, the dataset includes the numbers of confirmed COVID-19 cases and deaths, per country per date. Table 3 (p. 13) summarises the fields from OxCGRT that were included in *combinedCovidData*.

Table 3: Fields from the dataset *COVID-19 Government Response Tracker* included in *combinedCovidData*.

Field	Description	Field	Description
C1	Schools closing	H2	Testing policy
C2	Workplace closing	H3	Contact tracing
C3	Cancel public events	H4	Healthcare emergency investment
C4	Restrictions on gatherings	H5	Vaccine investments
C5	Close public transport	H6	Facial coverings
C6	Stay at home requirements	H7	Vaccination policy
C7	Internal movement restrictions	H8	Protection of the elderly
C8	International travel controls	chIndex	Containment and health index: C1-8, H1-4, H6-8
E1	Income support	esIndex	Economic support index: E1-2
E2	Debt/contract relief	stIndex	Stringency index: C1-8, H1
E3	Fiscal measures	grIndex	Government response index: C1-8, E1-2, H1-3, H6-8
E4	International support	Confirmed cases	
H1	Public information campaigns	Confirmed deaths	

Three of the fields have been of upmost importance for the present work. The first two are the number of cases and number of deaths. They were used to calculate the mortality rates, confirmed deaths per confirmed cases, recorded per country per date. Lastly, the Stringency index was prioritised over the other indices due to its exclusion of policies of lesser relevance for mortality mitigation, such as the economical and responsive (rather than preparatory) indicators. Figure 2 illustrates the mean stringency for the 186 countries the dataset covers.

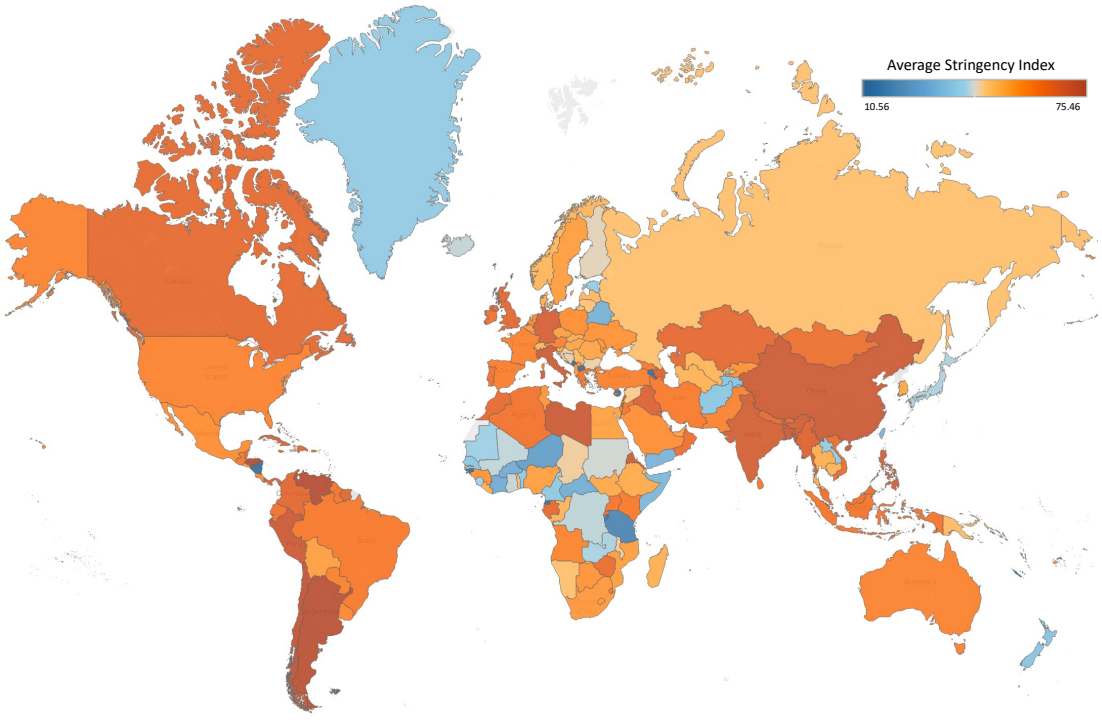

Figure 2: World map coloured by the average Stringency index, indicating relative restriction measurement levels on a global scale.

In addition to including only select columns from the OxCGRT dataset, some rows were also excluded from *combinedCovidData*. This is due to the original dataset often containing several daily entries for some countries, for which restrictions were registered both on national and regional levels. As the present work only regards national policy implementations, the indicator flags described in Table 8 (Appendix A.4, p. 42) were used to extract only the entries that regarded policy implementations on a national level.

3.1.5 Country demographics

The final fields included in *combinedCovidData* are information on each country's demographics, fetched from the website Worldometers [16]. The dataset contains population sizes, mean ages, and how many percent of the populations are considered urban. By including these fields in the dataset, other values can be standardised by population size. Fields like testing availability or social distancing restrictions can be considered based on urbanisation statuses, while mortality rates can be compared to the population mean age. These data have also been relevant during the categorisation of countries, as explained in Section 2.4 (p. 9).

3.1.6 Master dataset assembly

The five data sources described in the previous sections were combined using Tableu Prep Builder (ver. 2021.3) [21]. The flow of the assembly is shown in Figure 3.

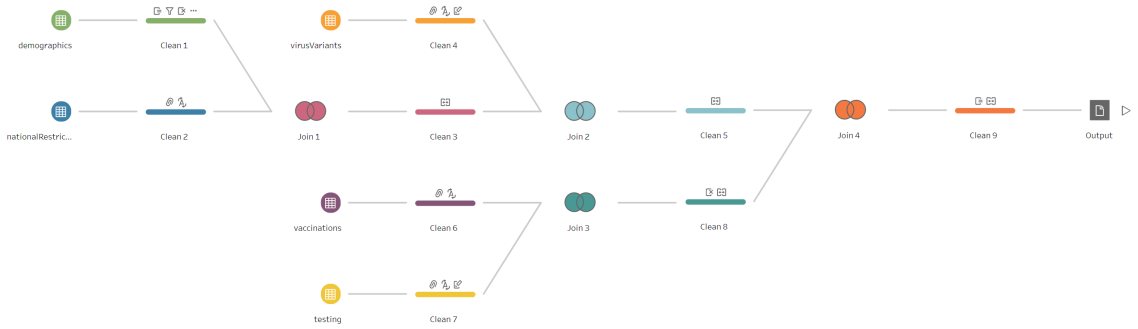

Figure 3: Tableau Prep Builder flow for combining the five datasheets into one output.

After importing the individual datasets, all were cleaned: unnecessary columns were removed; the fields *country* and *date* were assigned the roles *Country/Region* and *Date*, respectively; and field names were adjusted as to not overlap with similar fields from other sheets (e.g. *source* and *url*). The datasets were then merged using Country and Date as join clauses, except for *demographics* which were only joined using Country. After each joining, the resulting datasets were cleaned. These cleaning steps were required to combine Country and Date, as no datasets overlapped exactly for these fields. Before retrieving the final output, countries without any attached Date fields were removed, as these were only covered by the *demographics* dataset. Further, any unresolved country names were changed to ensure consistency. Examples include the reformulation of "Cote d'Ivorie" to "Ivory Coast", and "Curacao" to "Curaçao". The final output was exported to Excel as *combinedCovidData*, where the columns were ordered and colour-coded according to their origin dataset.

3.2 Analysis methods and tools

3.2.1 Shifting mortality

When analysing the *combinedCovidData* dataset, the data on mortality was shifted forward in time relative to the remaining fields. This is justified by considering the time between an infection and the potential death of the infected. For instance, an individual is infected during a period with low restriction stringency. Before any symptoms emerge, the country implements strict policies to mitigate transmission. Some time later, the individual develops severe symptoms and dies from the disease, upon which the mortality rate increases. Had this updated mortality rate been compared to the date of death, it would seem like the strict policies in place at the time were ineffective, seeing as the mortality rate still increased. However, the individual was infected previous to their implementation. Rather, it is the day of infection that the stringency level should have been recorded for comparison with the updated mortality rate.

Baud et al. estimated that the mean period between SARS-CoV-2 exposure and possible COVID-19-related death is 22 days. This includes an incubation period of 12 days, and 10 days from symptom onset until admission into the intensive care unit (ICU). The present analysis utilises a shift of 21 days, as a shift of three weeks were considered as more feasible for the analysis. The shift in mortality data offers clearer insights into the effects of the governmental policies implemented.

The shift in mortality was conducted by replacing the mortality data on day $X + 21$ by data on day X, for each country. The data on day X is then set to an empty numeric value, thus effectively deleted. The outcome is similar for the final 21 mortality entries for each country. These would have been assigned to dates in the future, for which no restriction data is recorded.

3.2.2 Initial visualisation, Shapiro-Wilk test, and Spearman correlation analysis

To gain an initial impression of the collected data, Tableau Desktop Public Edition (ver. 2021.3) [23] was used to plot various fields from *combinedCovidData* against date for all and for subsets of countries. Examples of fields include the OxCGRT indices, mortality rates, number of COVID-19 cases, and virus variant abundances. This visualisation aided in the determination of the project aim.

The fields of mortality and restriction stringency were viewed by plotting the mean values per day for both fields. The resulting graph was used to explore a possible correlation between the two by visual interpretation. A Shapiro-Wilk test (described in Appendix B.2, p. 43) was conducted to test whether the combined mortality and stringency were normally distributed for all countries. The test investigates the null hypothesis that a given sample comes from a normally distributed population, and a p-value < 0.05 indicates that there is a significant difference between the sample and a randomly generated sample of normally distributed variables. In turn, this indicates that the sample is not normally distributed. It is important to run a test on normality before performing further analysis, as many statistical tests, like T-test and ANOVA, rely on a normally distributed sample.

Using the built-in function *shapiro.test()* in R, for each country on both their mortality and stringency data allows one to run such an analysis on the distribution of the samples. The Shapiro-Wilk test showed that the sample for both mortality and stringency was significantly different from that of a randomly generated normal distribution sample, indicating that the samples are not normally distributed. For more detail on how this analysis was performed, see the R-script in the GitHub (Appendix C, p. 45).

A Spearman rank correlation was chosen to investigate the relationship between the two samples, with the aim of confirming the connection between stringency and mortality. As the Spearman rank correlation does not assume linearity nor normally distributed samples, it is appropriate for the utilised data. The test outputs a measure of statistical dependency between the ranking of two variables, which it uses to assess how the relationship between the variables can be described. The theoretical background behind the Spearman analysis is provided in Appendix B.1 (p. 43).

By using the built-in *.cor_test()* function in R, with the variable method set to Spearman, it is possible to retrieve a measurement of whether the relationship between the variables mortality and stringency index is monotonic or not, for all countries and dates present in the dataset. For more detail on how this analysis was performed, see the R-script in the GitHub (Appendix C, p. 45).

3.2.3 Narrowing the scope: selecting groups of countries and their representatives

Through a brainstorming process within the team, potential groups of countries were suggested. Their formation were based on their geographic locations, previous experience with SARS viruses, and the extent of the initial wave of the SARS-CoV-2 pandemic. Section 2.4 (p. 9) details the categorisations.

The daily mortality rates and stringency indices were plotted in Tableau for each of the chosen countries within a group to discover their believed similar trends. Supporting the visual interpret-

ation, boxplots an a pairwise Wilcoxon signed-rank test were generated in R for each group.

The Wilcoxon rank sum test is a non-parametric statistical hypothesis test used to calculate a pairwise comparison between group levels with corrections for multiple testing. The Wilcoxon rank sum test was favoured over other statistical analyses due to the samples not being normally distributed (see Section 3.2.2, p. 15). On a wide variety of data sets, the Wilcoxon test has a greater statistical power than the T-test, and is more likely to produce a statistically significant result for non-normally distributed data. For more background on the Wilcoxon analysis, see Appendix B.3 (p. 43). The R-function *pairwise.wilcox.test()* was used to conduct the analyses, with the input being the grouped countries and their mortality or stringency. Bonferroni was chosen for p-value adjustment to explicitly avoid type I errors: rejecting the null hypothesis even though it is accurate [24]. For more details on how this analysis was conducted, see the R-script in the GitHub (Appendix C, p. 45).

The choice of one representative country from each of the five groups were first conducted based on visual determinations. Since only rough statements can be made on the basis of these representations, the results from the Wilcoxon rank sum tests are the main decision features. The statistical significance of such correlation is indicated by a p-value with $p \geq 0.05$ being necessary to show a noteworthy correlation. If a country is part of several significantly valued pairings, then it is considered a potential representative. If this is true for multiple countries, the distribution of stringency and mortality data is utilised to make the decision on the representative country. In this case, it is advantageous if the distribution of the data is small, allowing for a simpler identification of trends in the final analysis. In situations were the test nor boxplot yielded a clear candidate, subjective factors may have impacted choice of country.

3.2.4 Breaking down the fields impacting mortality for each representative country

As the group decided to consider multiple policies and restrictions as an explanation for mortality, a correlation matrix was calculated. It was generated using R's built-in function *.cor()*, with Spearman as the specified method. From the resulting matrix, the column of ρ values representing the relationships between mortality and the various COVID-19 measurements was extracted. The results were visualised with bar charts. See Appendix C (p. 45) for the R-script used to create the correlation matrix.

4 Analysis

4.1 Correlation between restriction stringency and COVID-19 mortality rate

To understand the overall connection of the stringency index and the mortality rate from COVID-19, the mean over all 145 countries for which data was recorded, is plotted for both the index and the mortality rate. The data is presented in Figure 4 and gives a first rough estimate of trends. Keep in mind that the figure uses dual axes, where the scale is 0-100 for stringency and 0-1 for mortality to more clearly visualise the trend. Firstly, it can be noted that the overall relative trend of both the stringency index and the mortality are similar over the whole time frame. Secondly, it is worth to note the rapid increase in both mortality and stringency index right at the beginning of the pandemic. This results in a major peak for both variables after which it subsequently decreases until reaching a steady level with minor deviations with further progression of time. Having described the overall trend in the data, there are noteworthy divergences in certain areas.

Firstly, the early spikes in mortality, with the largest being the mean mortality of February 19$^{\text{th}}$ 2020. Considering the standard deviation, represented by the colour intensities of the graphs, this region is where the mortality varies the most between the countries. This is explained by the fact that only 33 countries were reporting mortality data at this early stage in the pandemic; one of which is Iran with a mortality of 1.0 (both confirmed cases resulted in deaths), and France

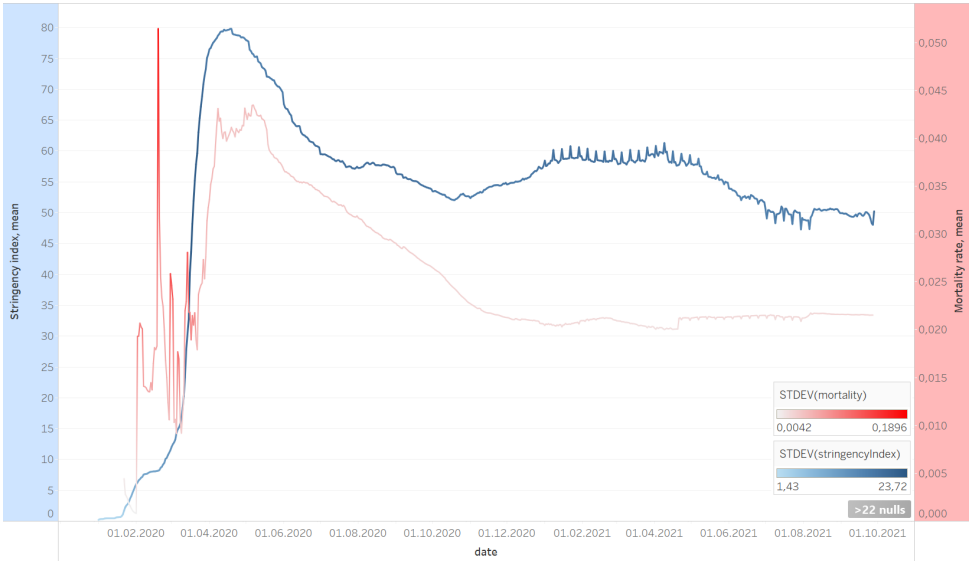

Figure 4: Worldwide mean of stringency index and mortality rate from the 1st of October 2020 to the 28th of September 2021: The mean stringency index over 145 countries (blue) is displayed on a scale between 0 and 100, while the mean mortality for 89 countries (red) is displayed as a relative value (0-1). The darkness of the curves indicate the magnitude of standard deviation for each value.

with mortality rate 0.083 (one death in 12 cases). The mortality rate in Iran the next day is reported as 0.4 (three new cases registered, and no new deaths). In contrast, there are over 180 countries reporting their mortality data one year later. On February 19$^{\text{th}}$ 2021, Iran and France report mortality rates of 0.0381 (59,341 deaths in 1,558,159 cases) and 0.0236 (8,3542 deaths in 3,541,282 cases), respectively. Thus the large deviations and fluctuating mean mortality rates can be attributed to the low number of countries reporting data at this stage, and the higher tendency of these cases to have a negative outcome.

Further of note are local peaks appearing both in the mean stringency index data and mortality rate. These appear roughly after December 2020 and are most prominently seen in the curve for stringency (see Figure 4). Reasons for the appearance of these peaks could be manifold. One reason could be, considering the vast amount of countries, that governmental policies started greatly varying in a short time frame, thus resulting in the aforementioned peaks. Additionally for the mortality it is possible that certain data was not recorded properly and that data is missing, resulting in the peaks. It also has to be considered that faulty data acquisition is possible. Despite these possibilities, the data still proves valuable for the subsequent analysis, as countries are analysed individually. Thus, as long as unexpected deviations do not reappear in the data for single countries, the peaks seen here are not of further importance.

In addition, a Spearman's rank-order correlation was computed to determine the relationship between mortality and the stringency index across all countries for the whole time period, for which data is available. There was a intermediate, positive correlation between these measurements, which was statistically significant ($r(117766) = 0.3397174$, $p = 2.2\text{e}{-16}$). As the test was run on data across all countries, it is not surprising that it reveals a positive relationship between mortality and stringency, meaning that they increase or decrease in cohesion. As an increasing stringency is often implemented as the mortality increases and then lifted again as mortality decreases, this relationship does make sense for the overall world. However, this illustrates that conducting further analysis on all of the worldwide data is not useful.

4.2 Determining groups of countries

As previously discussed the analysis of all countries did not seem beneficial. It was therefore decided to group countries by certain general trends. The means by which the countries were grouped are discussed in Section 2.4. The groupings of the countries is further discussed in Section 3.2.3. To give a quick overview of the groups analysed: The first group consists of countries that had an overall similar initial rapid progression of COVID-19, such as Italy and France. The geographic location and the overall high social capital, which distinguishes these countries from most, were the reasons for the formation of group 2, which consists of some Nordic countries. The third group consists of countries which had experience with the previous SARS outbreak in 2002-2004. It can be therefore assumed, that infrastructure and experience for dealing with respiratory diseases already exists. The third group summarises countries in central Europe with an overall lower mortality compared to the countries in the first mentioned group. The fourth and final group also consists of countries in central Europe with the difference being, that the mortality was overall higher.

For each of these countries are representative is decided by analysing the results from Wilcoxon tests (see Section 3.2.3). The conclusions from these tests and further comments on the groups are presented in the following sections.

4.2.1 Heavily impacted countries in Europe: Italy, France, Spain, and UK

The first group which is to be discussed, is formed by countries that had a devastating peak in mortality in the beginning of the pandemic. Here the countries chosen are Italy, France, Spain and UK. While it is possible, that other countries showed similar behaviour in mortality, the analysis is restricted to these countries, as these were, next to being few of the first countries hit by COVID-19 in Europe, the ones which were most prominent in international discussions when it came to the effectiveness of lockdown measurements.

In Figure 5 the mortality rate and stringency index are plotted for the four countries. It can be seen, that the progression of the mortality rate follows an overall dome like shape before peaking around June/July 2020 and then decreasing. The mortality then reaches an almost steady value for all countries beginning around December of 2020. The highest value for mortality is recorded in France with 14 %. Likewise, similar trends in-between countries can be observed in the progression of the Stringency Index. The stringency index sharply increases in the beginning of 2020 with first cases appearing before sharply decreasing with the beginning of May 2020. A second peak in the stringency index between November 2020 and June/July 2021 can be observed as well. But it can be noted here, that during this time period, the magnitude of the stringency index varies greatly between countries. This is due to the different approaches then chosen by the governments to handle the resurging number of cases. A prominent example would be the UK, where governmental restrictions impacting freedom of movement dropped sooner due to the rapid progress in the countries vaccination program. It is of note however, that during this second peak in the stringency index, no correlation peak in the mortality data can be observed.

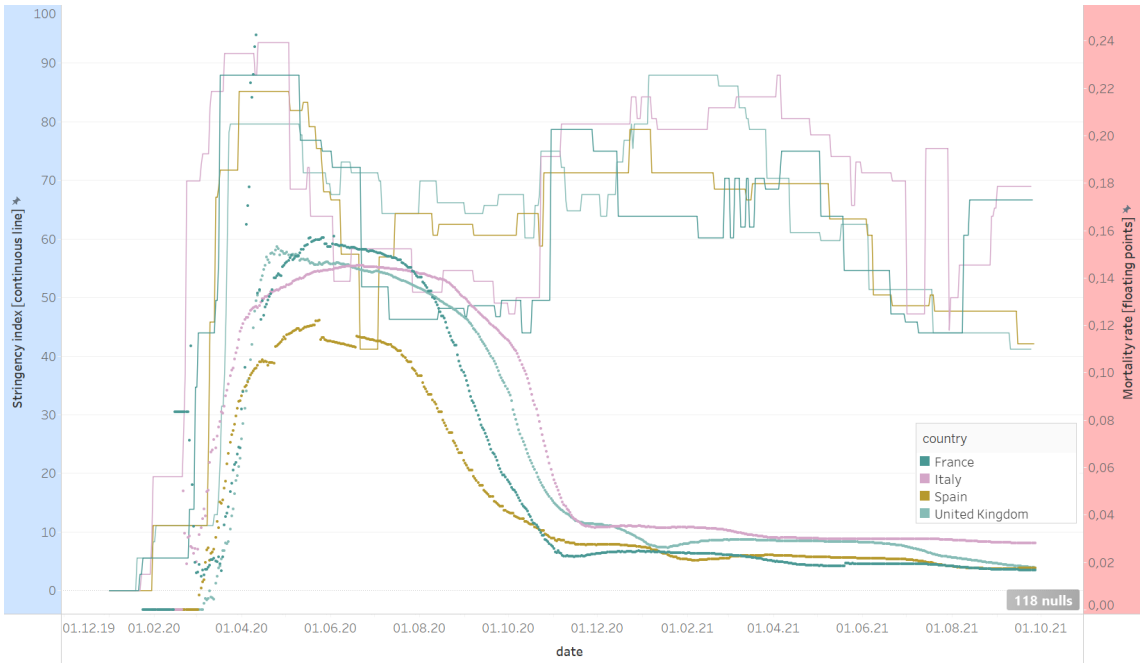

Figure 5: Development of stringency index and mortality rate for the countries being heavily impacted at the beginning of the pandemic. The countries depicted here are France (green), Italy (purple), Spain (brown) and the UK (light blue). The solid lines in the plot represent the stringency index while the dotted curves show the progression of the mortality rate. The stringency index is plotted as percentage (0-100), while the mortality rate is plotted as a numerical ordinal scale (0-1).

As discussed in Section 3.2.3 pairwise Wilcoxon rank sum tests were utilised to determine a representative for the group, by first identifying which countries are most similar in progression to each other within the group. The method is chosen due to the data on mortality and stringency not being normally distributed. The results from the tests are visualised in Figure 6. Additionally to the tabulated values from the Wilcoxon test, boxplots are used to represent the behaviour of the mortality.

As mentioned, a Wilcoxon test checks the following null hypothesis:

$$H_0 : \text{The difference in group means is zero}$$

As discussed in Section B.3 the null hypothesis is considered to be rejected, if $p \leq 0.05$. In Figure 6 it can be seen that only for France and Spain the result from the pairwise Wilcoxon test does not justify a rejection of the null hypothesis both for the mortality ($p = 1$) and stringency ($p = 0.1167$). Therefore it can be concluded that the that the progression in those two countries can be considered to be similar, while all other possible country pairings exhibit a significant deviant behaviour.

To further investigate which country should be chosen as a representative, a boxplot was drawn for each country for both mortality and stringency values. From the boxplot, it can be clearly seen that Spain has an intermediate median value in both mortality and stringency, as well as a lower variation in mortality than the rest of the representatives. Due to the lower variations, trends and tendencies in the data are more easily identified, and Spain is therefore chosen as the representative for the initial spike countries.

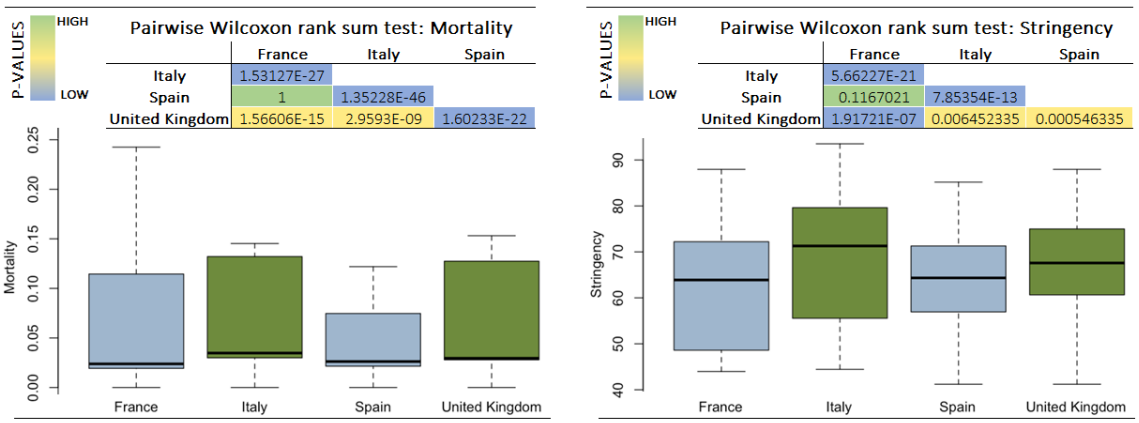

Figure 6: Results from the Wilcoxon tests for mortality (left) and stringency index (right) with boxplots illustraing the behaviour of the examined values for the countries showing an early spike in cases: For both mortality and stringency index the results from Wilcoxon test are summarized on the top in tabular form. Here the colour indicates the relative magnitude of the p-values with blue representing low and and green representing high p-values. The boxplots shows the behaviour of mortality (left) and stringency (right) for each country individually. In each box the bold black line represents the median of the data for the variable while the top and bottom rule of the boxes represent the 25- and 75-percentiles respectively. The maximum and minimum values are represented by whiskers.

4.2.2 Nordic countries: Denmark, Finland, and Norway.

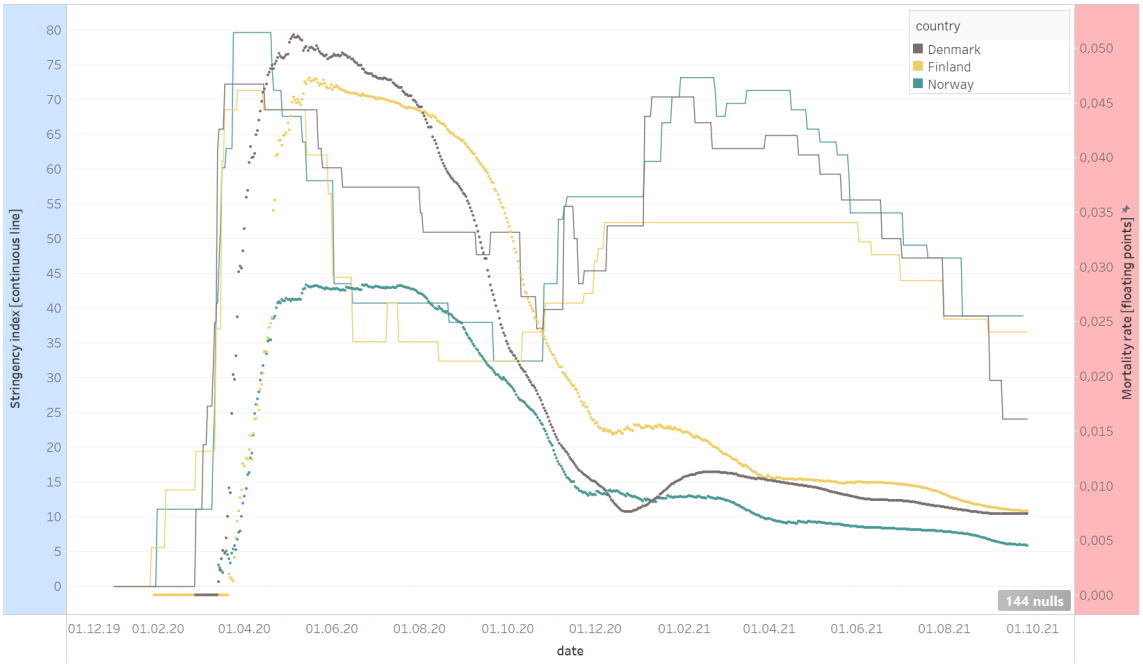

Figure 7: Development of stringency index and mortality rate for the Nordic countries. The countries depicted here are Denmark (brown), Finland (yellow), and Norway (green). The solid lines in the plot represent the stringency index while the dotted curves show the progression of the mortality rate. The stringency index is plotted as percentage (0-100), while the mortality rate is plotted as a numerical ordinal scale (0-1).

As mentioned previously: the second group comprises of a selection of Nordic countries, mainly Denmark, Finland and Norway. The development of the stringency index and mortality for these countries can be seen in Figure 7.

It can be observed in Figure 7, that, similarly to the initial spike countries, a first peak in both stringency and mortality is visible for the Nordic countries as well. In comparison to the mortality peaks of the early spike countries, here the magnitude of the peak is much lower. While France had a mortality peak at 14 % for the Nordic countries have their highest mortality in Denmark with roughly 5 %. This has to be considered, when comparing the figures as the range of scale for the Nordic countries is much smaller than for the early spike countries.

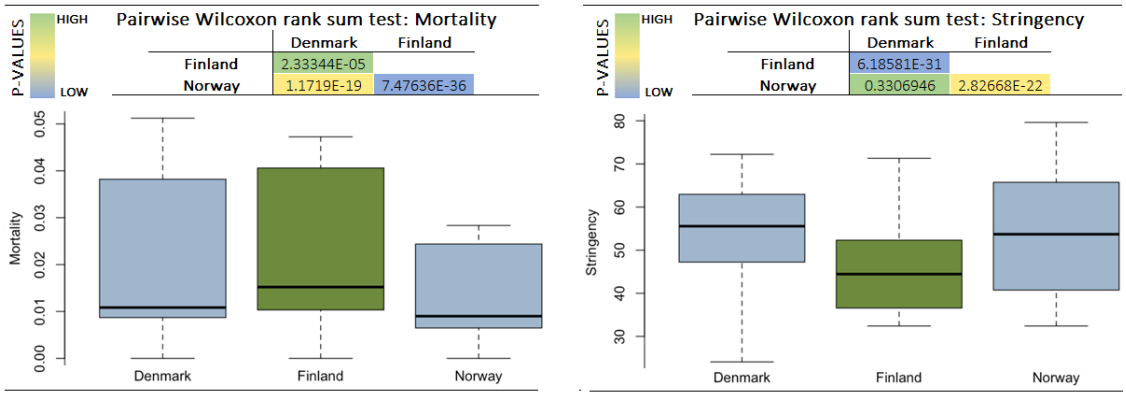

Figure 8: Results from the Wilcoxon tests for mortality (left) and stringency index (right) with boxplots illustrating the behaviour of the examined values for the countries showing an early spike in cases: For both mortality and stringency index the results from Wilcoxon test are summarised on the top in tabular form. Here the colour indicates the relative magnitude of the p-values with blue representing low and and green representing high p-values. The boxplots shows the behaviour of mortality (left) and stringency (right) for each country individually. In each box the bold black line represents the median of the data for the variable while the top and bottom rule of the boxes represent the 25- and 75-percentiles respectively. The maximum and minimum values are represented by whiskers.

As discussed in Section 3.2.3, Wilcoxon test is used to identify the representative of this group. The results can be seen in Figure 8. It can be seen that for mortality the null hypothesis is rejected for all country pairings. However it can be noted, that the pairing of Denmark with Finland has the highest p-value, despite this, the null hypothesis is still rejected. On the other hand, for the stringency index, the null hypothesis is accepted for the Denmark-Norway pairing with $p \approx 0.33$. Therefore a high similarity in stringency index progression can be assumed between Denmark and Norway.

Looking at the boxplots representing the distribution of mortality data (see Figure 7) it can be seen, that the variation within the data is lowest for Norway, with Denmark and Finland showing much wider distributions as well as having greater maximum values. The mean value is similar for all three countries, only being slightly higher for Finland. In the boxplots for the stringency index it can be seen, that here Finland shows the smallest variation, while Norway shows the greatest. Special attention can be paid to the greatly varying maximum and minimum value from the quantiles for Denmark. The mean stringency value is similar between Denmark and Norway while being significantly smaller for Finland. Due to Finland differing to greatly when either comparing the country with either Norway or Denmark, Finland is excluded as a representative country for the group. When trying to decide between Denmark and Norway, the high p-value for the stringency index between Denmark and Norway is the main indicator. Further taking into considerations the results from the different pairings and the distribution in mortality and stringency data, Norway proves to be the best candidate to be the representative country for the group. Therefore, Norway is chosen to represent the Nordic countries in future analysis in this report.

4.2.3 SARS-experienced countries: Hong Kong, Malaysia, South Korea, and Thailand

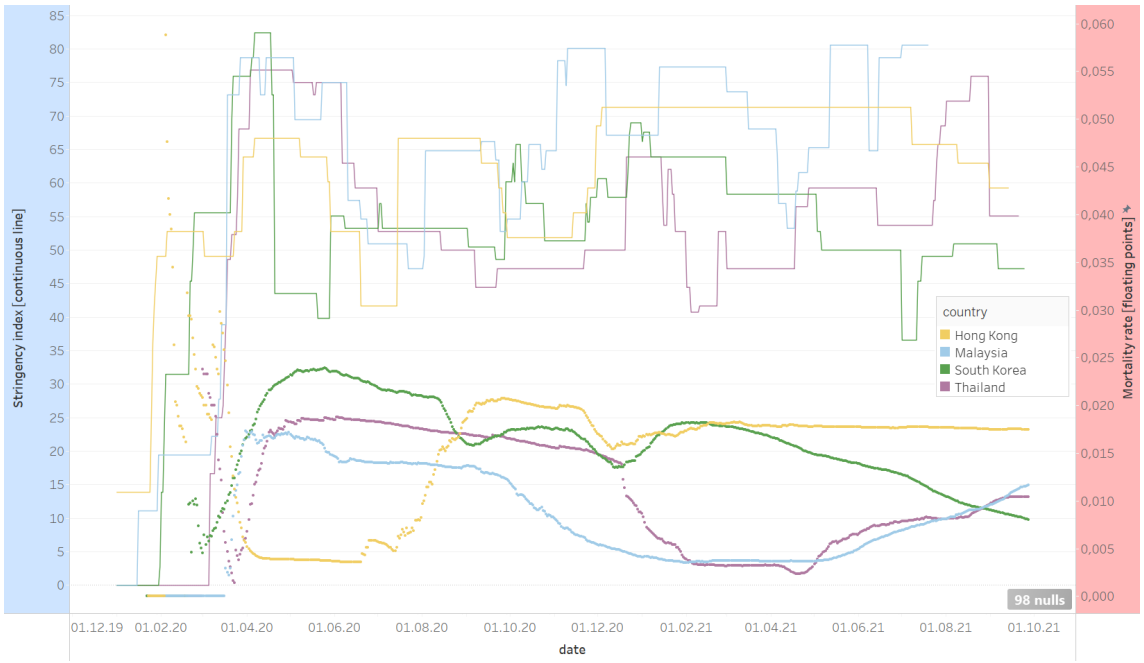

Figure 9: Development of stringency index and mortality rate for countries with previous experience with SARS. The countries depicted here are Hong Kong (yellow), Malaysia (light blue), South Korea (green) and Thailand (purple). The solid lines in the plot represent the stringency index while the dotted curves show the progression of the mortality rate. The stringency index is plotted on as percentage (0-100), while the mortality rate is plotted as a numerical ordinal scale (0-1)

This group consists of countries that were impacted significantly during the SARS-pandemic in 2002/3 (see Section 2.4.3). The additional experience in dealing with pandemics of this type is expected to result in a significant difference in the course compared to European countries. Before investigating that, however, it is again necessary to designate a representative country for this group. Similar to before the progression in stringency index and mortality rate are visualised in Figure 9. At first glance, the absence of a clear maximum in the mortality rate at the beginning of the pandemic immediately stands out. Also, the difference in mortality rates in between countries is significant, with South Korea showing a higher mortality rate on average than the other three countries. Furthermore it can be noted, that for Thailand, South Korea and Hong Kong the mortality rate stays stagnant for long amounts of time. Additionally, Hong Kong does not show any resemblance of a maximum in mortality at the beginning of the pandemic, one only appearing in September 2020. The progression of the stringency however are similar. Here a clear peak during the beginning of the pandemic can be distinguished. After that a overall high level of stringency is retained for all countries, but differences in approaches are clearly similar, with Malaysia maintaining a higher stringency index on average compared to the other countries.

As correlations in-between countries are not easily determined from Figure 9 pairwise Wilcoxon tests are applied once more to determine a representative. The results can be seen in Figure 10. They can be summarised as follows: For mortality only between Hong Kong and South Korea the null hypothesis is accepted with $p = 1 \geq 0.05$, while for the stringency index the test only approves the pairing of South Korea and Thailand with $p \approx 0.83 \geq 0.05$. After this analysis, South Korea seems to be the most likely candidate for representing this group. Further taking into account the distribution of the mortality and stringency data, South Korea seems to have small distributions in both variables, similar to Hong Kong, but with significantly fewer outliers. It can however be seen clearly here, that South Korea indeed had the highest mortality rate, as well one of the highest mean mortality values, with only Hong Kong showing a higher mean value. Further, it can be noted, that Malaysia seems to show a vastly different behaviour from the other countries as expected. Thus implying, that further differentiation might be necessary when grouping countries.

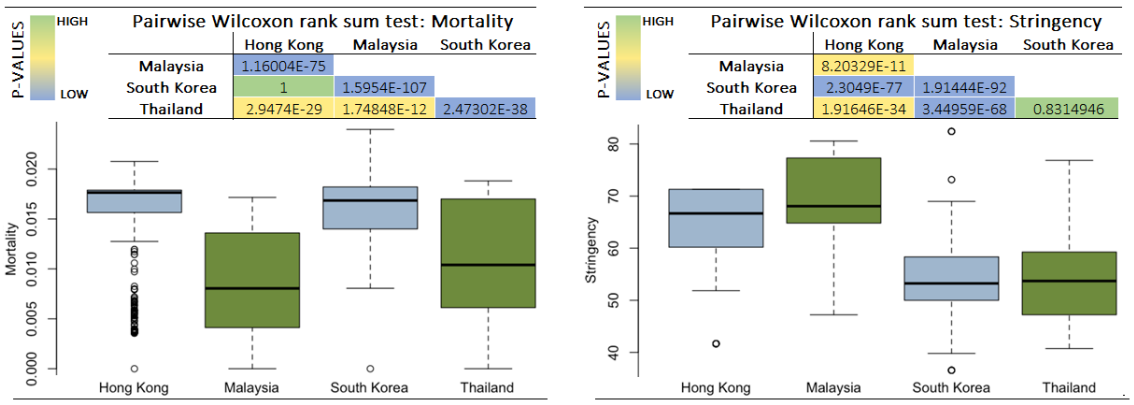

Figure 10: Results from the Wilcoxon tests for mortality (left) and stringency index (right) with boxplots illustrating the behaviour of the examined values for the countries showing an early spike in cases: For both mortality and stringency index the results from Wilcoxon test are summarised on the top in tabular form. Here the colour indicates the relative magnitude of the p-values with blue representing low and and green representing high p-values. The boxplots shows the behaviour of mortality (left) and stringency (right) for each country individually. In each box the bold black line represents the median of the data for the variable while the top and bottom rule of the boxes represent the 25- and 75-percentiles respectively. The maximum and minimum values are represented by whiskers.

Due to having the most approved country pairings by the Wilcoxon test and overall narrow distributions in data, South Korea is chosen as a representative.

4.2.4 Central Europe with lower mortality: Austria, Czechia, Germany, and Poland

The fourth group consists of a selection of central European countries that have a comparatively lower overall mortality rate compared to the countries in Section 4.2.1. In Figure 11 the development of stringency in mortality is summarised. It can be seen, the Czech Republic has the lowest maximal value for mortality rate, while Poland has the highest. Furthermore it can be observed, that after the first peak, the mortality rate dips before rising again and then stagnating. This occurs for all four countries to a certain degree. In difference to the other three countries, Germany seems to have a wider peak in the mortality rate. Similar to before it can be observed that the stringency index peaks early in the pandemic before falling and the rising again beginning in October 2020.

Utilizing Wilcoxon tests to determine a representative for the central European countries results in following country pairings where the hypothesis is approved for mortality (see Figure 11): Austria-Czechia and Germany-Poland. In case of the stringency index, only the pairing containing Austria and Poland leads to an acceptance of the hypothesis. These observations would indicate that Austria would be a valid choice as a representative. Here a further consideration comes into effect, not elaborated previously. Due to the part of the team having a close connection to Germany and having means to access native news sources, Germany is chosen as a representative over Austria. The results from the Wilcoxon test only partially support this decision. Nonetheless, from the boxplots it can be seen, that the data for Germany on mortality and stringency contains no outliers, unlike Czechia and Poland, and shows the smallest distribution in the stringency index. Therefore, whilst only being the second choice due to the results from the Wilcoxon tests, the boxplots clearly show, that Germany is to be preferred as a representative. Taking into account the project teams experience, Germany is chosen as a suitable representative for the central European countries.

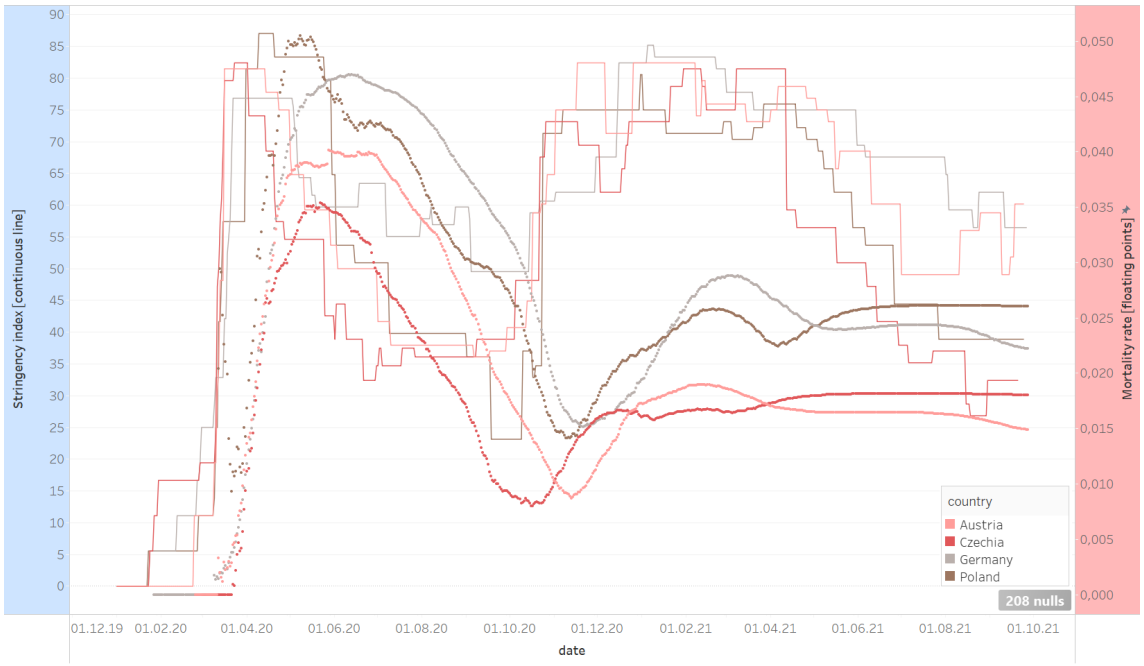

Figure 11: Development of stringency index and mortality rate for central European countries. The countries depicted here are Austria (orange), Czechia (red), Germany (grey) and Poland (brown). The solid lines in the plot represent the stringency index while the dotted curves show the progression of the mortality rate. The stringency index is plotted as a percentage (0-100), while the mortality rate is plotted as a numerical ordinal scale (0-1).

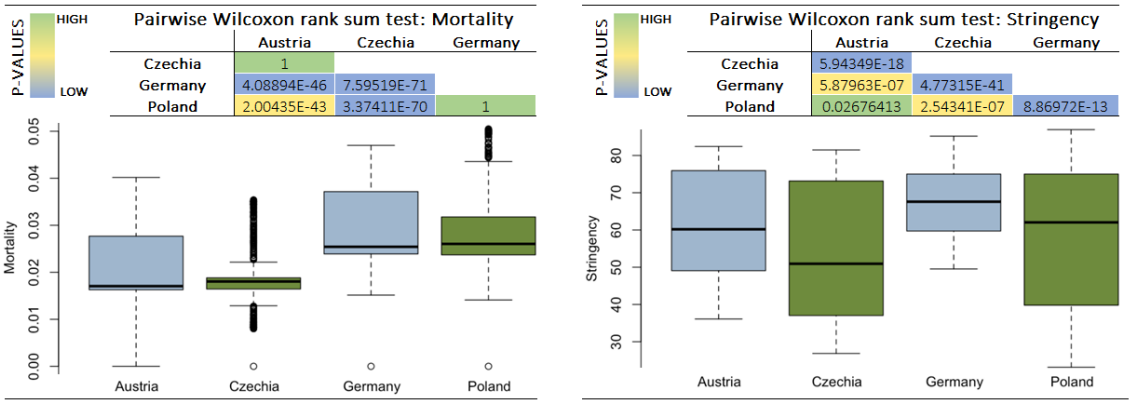

Figure 12: Results from the Wilcoxon tests for mortality (left) and stringency index (right) with boxplots illustrating the behaviour of the examined values for the countries showing an early spike in cases: For both mortality and stringency index the results from Wilcoxon test are summarised on the top in tabular form. Here the colour indicates the relative magnitude of the p-values with blue representing low and and green representing high p-values. The boxplots shows the behaviour of mortality (left) and stringency (right) for each country individually. In each box the bold black line represents the median of the data for the variable while the top and bottom rule of the boxes represent the 25- and 75-percentiles respectively. The maximum and minimum values are represented by whiskers.

4.2.5 Central Europe with higher mortality: Belgium, Hungary, and The Netherlands

The final group is composed of some remaining central European countries, that, compared to the fourth group (see Section 4.2.4), have a higher peak and average mortality. It contains Bel-

gium, Hungary and the Netherlands. Further details on the choice of country can be found in Section 2.4.4.

The visualisation of the progression of mortality and stringency for these countries can be found in Figure 13. For each country the by now familiar peak in mortality in the first half of the time period can be seen for all three countries with Belgium reaching the highest mortality rate with roughly 16 %. After the maximum the mortality rate drops drastically over the following months before stagnating between 1 and 3.8 % with the highest rate belonging to Hungary. The lowest value in this range belongs to the Netherlands which also show the lowest maximum in mortality rate of all three countries. For the stringency index it can be observed, that similar to all previously discussed countries, there is a peak corresponding to the rise in mortality at the beginning of the time period. The index than stagnates for roughly one to two months before dropping to around 60. After that, there is a great deviation between countries in stringency index reflecting the different approaches chosen by the governments in the subsequent handling of the pandemic.

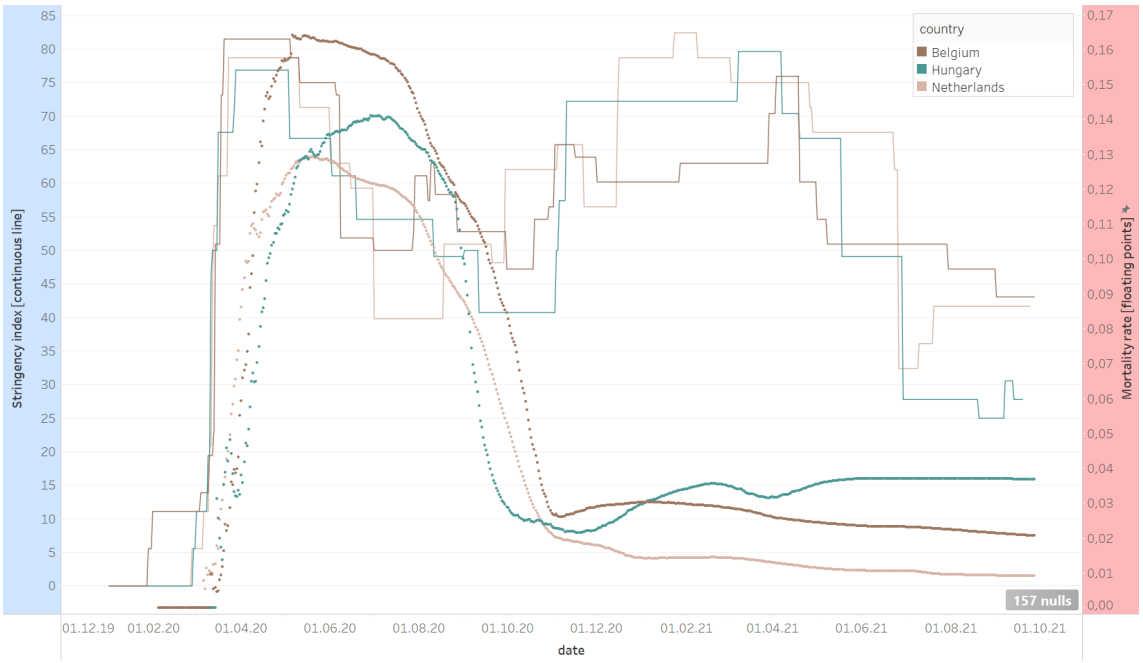

Figure 13: Development of stringency index and mortality rate for central European countries with higher mortality rate. The countries depicted here are Belgium (brown), Hungary (green) and the Netherlands (yellow). The solid lines in the plot represent the stringency index while the dotted curves show the progression of the mortality rate. The stringency index is plotted as percentage (0-100), while the mortality rate is plotted as a numerical ordinal scale (0-1).

As before, the same procedure is applied to determine a representative for this group. The results form the Wilcoxon tests can be found in Figure 14. It can be seen from the table that the null hypothesis is rejected for all pairings when analysing the mortality data. Same can be noticed for the stringency data, with the null hypothesis only being accepted for the pairing containing Belgium and Hungary. While the null hypothesis is rejected for all pairings in case of the mortality, the low values for the Belgium-Hungary and Hungary-Netherlands pairings can be noted. Likewise the medium p-value for the stringency data for the Belgium-Netherlands pairing can be noted. Considering the results from the Wilcoxon tests for the Stringency index, either Belgium and Hungary seem suitable as a representative for the group. Further analysing the variation in data using the boxplots in Figure 14 it can further be noted, that the variation in mortality data is smallest for Hungary and largest for Belgium. For the stringency data the variation is smallest for Belgium while very similar between Hungary and the Netherlands. These observations seem to fall in favour of Belgium being chosen as the representative country. To solidify this decision, the median for stringency and mortality can be taken into consideration. It can be seen, that the median for Belgium lies in-between the one for Hungary and the Netherlands, both for mortality

and stringency. Seeing that Belgium lies in range for both of the other countries and shares similarity in the box plots, further solidifies the choice of Belgium for the group discussed.

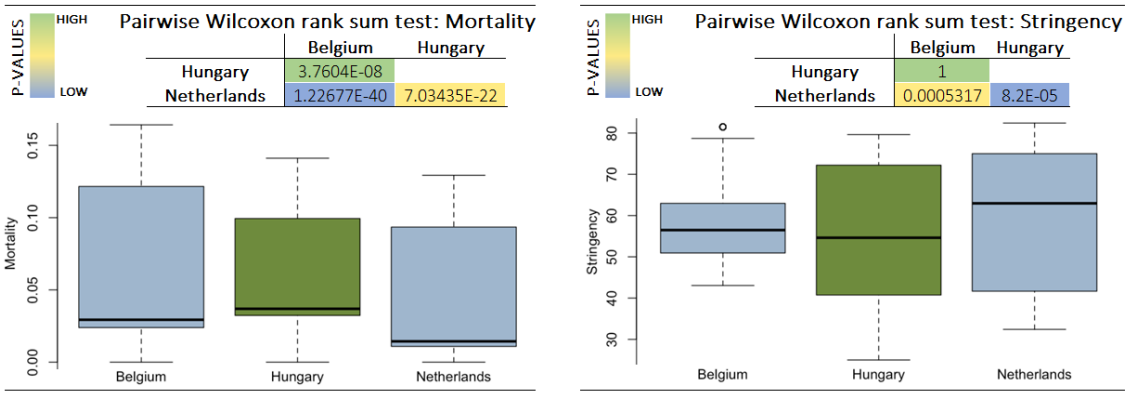

Figure 14: Results from the Wilcoxon tests for mortality (left) and stringency index (right) with boxplots illustrating the behaviour of the examined values for the countries showing an early spike in cases: For both mortality and stringency index the results from Wilcoxon test are summarised on the top in tabular form. Here the colour indicates the relative magnitude of the p-values with blue representing low and and green representing high p-values. The boxplots shows the behaviour of mortality (left) and stringency (right) for each country individually. In each box the bold black line represents the median of the data for the variable while the top and bottom rule of the boxes represent the 25- and 75-percentiles respectively. The maximum and minimum values are represented by whiskers.

4.3 Justification of mortality shift and implications for representative countries

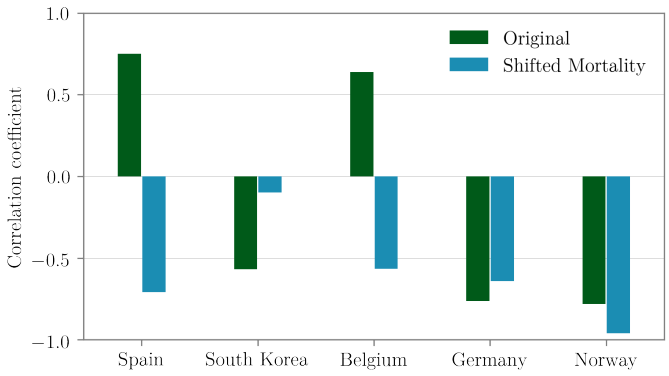

Figure 15: Visualisation of difference in correlation coefficients due to shifted mortality data in a restricted time period. The correlation coefficient is determined for the original mortality data (green) and the shifted mortality data (blue) for each country chosen as representative of their group. The correlation coefficients are calculated for a time period from the day $X - 1$, before measures where first lifted to 50 days after $(X + 50)$.

As discussed previously in Section 3.2.1 the mortality data was shifted in the following analysis to account for the time difference between infection and death from the same infection to further highlight the impact on governmental policies on the stringency index described in Section 3.1.4. From the curves presented in Figure 16 it can be seen, that there is usually one prominent peak in mortality and stringency index. The progression of the curves looks fairly similar for all countries in the beginning of the pandemic. First differences become visible when governmental policies are eased. It is therefore of special interest, how the mortality is effect by the falling stringency index after the date when measures are first lifted. As this date for the countries discussed is usually in

April or May of 2020 the vaccination progress does not affect the mortality yet. The time frame therefore gives the most insights into the effects of solely the governmental policies on mortality.

The following analysis, using the data from the time after policies are first lifted, shows the effects of shifting the mortality data as discussed in Section 3.2.1. For each of the representative countries the date on which the stringency index is first lower than the maximum value is identified as day X. For the complete analysis day $X - 1$ is taken into account to include the maximum stringency index value for reference. In the next step, the data for the subsequent 50 days is extracted and the correlation coefficient is calculated, using the Spearman method described in Section B.1. The duration of 50 days was chosen as it seems to be a reasonable time frame, in which the governmental policies can affect the infection. Additionally, this duration is not too heavily influenced by external factors.

The results for the representative countries are shown in Figure 15. For Spain and Belgium, it can be seen, that the sign and magnitude of the correlation coefficient change greatly when shifting the mortality data. The negative correlation coefficient implies that with governmental policies being lifted, the mortality increases, which is to be expected, assuming the policies are successful in promoting a decrease in infection. Therefore, shifting the mortality data makes the delayed effects of the policies visible. Unfortunately, contrary to exceptions, this effect cannot be seen for all representative countries. While for Norway, the magnitude of the correlation still increases, same cannot be observed for South Korea and Germany. For these two countries are formerly strong negative correlation on the original data set is decreased slightly, for Germany, and significantly, for South Korea. Trying to determine the reasons for this unexpected shift proves to be rather difficult. Nonetheless, shifting the mortality data seems to show a more accurate picture of the effects of the shift in stringency index. It was therefore conducted for the following analysis, unless stated otherwise.

4.4 Comparison of representative countries

The final part of this analysis focuses on the representative countries chosen in Section 4.2: Belgium, Germany, Norway, South Korea and Spain. The goal is identifying trends and correlations between mortality and stringency to give appropriate recommendations as discussed in Section 2.3.

For direct comparison the development of stringency and mortality rate are plotted in Figure 16. As the general trends have been already discussed in the previous sections, only some highlights are going to be discussed here. In the figure the especially high peak in mortality for Belgium stands out, with Spain having the second largest maximum. Lastly, the mortality for Germany is significantly higher than Norway and South Korea, that show a similar magnitude in mortality rate. Therefore, it is possible to assume a further sub-grouping of the countries into high and low mortality, with Spain and Belgium forming the high-mortality group. The stringency trends are difficult to interpret using the figure. It can only be observed, that two peaks are distinguishable. Due to the large variations in stringency it is unfeasible to make additional observations from the figure alone.

To gain further insights into the data and to be able to make appropriate recommendations, Spearman analysis was conducted. As the stringency index is composed of several ranked containment and closure policies (see Section 3.1.4), the value of each of these policies for lowering the mortality rate can be investigated. For the analysis, all governmental policies possible were investigated, not restricted to the ones used for the calculation of the stringency index (see Section 3.1.4). The results from the analysis are visualised in Figure 17. In the figure the correlation value ρ is reported. If the correlation between mortality and the investigated variable is statistically significant, the bar is highlighted in green. The correlation indicates the relation between two variables (see Section 4.1). In short, a positive correlation between mortality and e.g. restrictions on internal movement (C7) would imply, that a higher scrutiny has a higher mortality as a consequence. While this phrasing is partly exaggerated, as it seems highly unlikely, that less travel would promote more deaths, a positive correlation still indicates, that the measures does not seem to be effective. Therefore policies with a negative correlation are of most interest, as the negative value implies that for e.g. a tougher implementation of facial coverings (H6) the mortality rate falls. The strength of

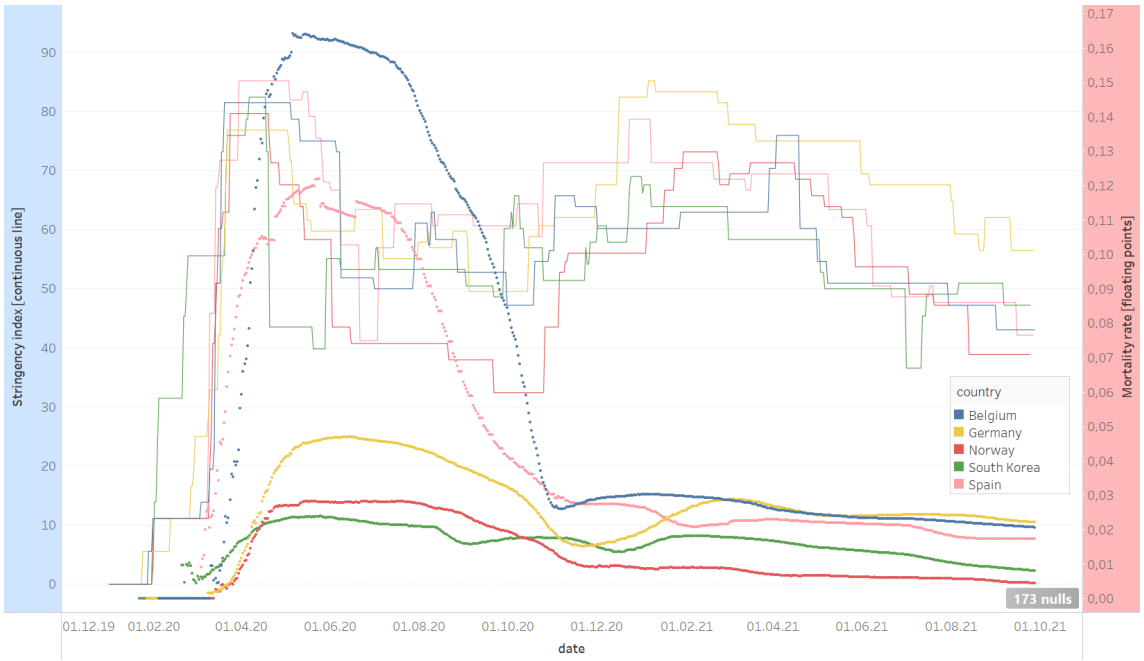

Figure 16: Development of stringency index and mortality rate for the chosen representative countries. The countries depicted here are Belgium (blue), Germany (yellow), Norway (red), South Korea (green) and Spain (orange). The solid lines in the plot represent the stringency index while the dotted curves show the progression of the mortality rate. The stringency index is plotted as percentage (0-100), while the mortality rate is plotted as a numerical ordinal scale (0-1)

correlation is indicated by the value ρ. While 0 generally refers to no correlation, values between 0.4 and 0.7 indicate a moderate correlation, positive or negative depending on the sign, between variables. For $\rho \geq 0.7$ the correlation is strong, with 1 being the maximum possible value implying a linear relationship between the variables. Values between 0.1 and 0.3 generally imply a weak correlation. These values are for guidance only and do not represent hard limits.

It can be observed in Figure 17 that there are certain policies which promote the decrease in mortality rate over all countries. One example would be the implementation of vaccination policies (H7). When looking at the countries individually, it can be seen, that H7 actually had the biggest negative correlation to the mortality rate in Belgium, compared to all other policies with testing policies (H2) ranking second. For Germany the stay at home requirements (C6) show to be effective with the closing of public transport (C5) following close behind. In Norway, compared to Belgium and Germany a greater amount of policies are negatively correlated to the mortality rate, with H7 showing the strongest negative correlation. For both South Korea and Spain, only few results exhibit the desired strong negative correlation. In case of both countries, the vaccination policies yield the strongest negative correlation.

It is not unexpected, that vaccination policies result in the strongest effect in reducing the mortality rate, as the effects of vaccinations reduce the number of deaths and lessen the severity of the disease. As it cannot be expected for vaccinations to be available at the beginning of a pandemic, due to a potentially unknown virus, it is necessary to explore the effects of the additional measures. As mentioned previously, in Belgium the testing policies (H2) proved highly valuable, as well as restrictions on social gatherings (C4). In Norway, C5 and C6 show strong negative correlations, as well as in Germany. In South Korea only restrictions on gatherings (C4) and implementation of vaccination policies (H7) result in a weak correlation to the mortality rate. For Spain, only C4, C6 and H2 show a weak negative correlation. It can be seen from this discussion, that there is an overlap to some extent, in effective implementations over all different countries. Another policy, which shows a negative correlation at varying extents, is the need for facial coverings (H6). This policy shows a strong correlation in e.g. Belgium and Norway and a significantly smaller, but still

A

Abbrev.	Description
C1	School closing
C2	Workplace closing
C3	Cancel public events
C4	Restrictions on gatherings
C5	Close public transport
C6	Stay at home requirements
C7	Restrictions on internal movement
C8	International travel controls
E1	Income support
E2	Debt/contract relief
E3	Fiscal measures
E4	International support
H1	Public information campaigns
H2	Testing policy
H3	Contact tracing

Abbrev.	Description
H4	Emergency investment in healthcare
H5	Investment in vaccines
H6	Facial coverings
H7	Vaccination policy
H8	Protection of elderly people
chIndex	Containment health index
esIndex	Economic support index
grIndex	Government response index
stIndex	stringencyIndex
#dailyVacc	numDailyVaccinationsPerMillion
#fullyVacc	numFullyVaccinatedPerHundred
#tests	numTestsPerThousand
alpha	alpha variant abundance
delta	delta variant abundance

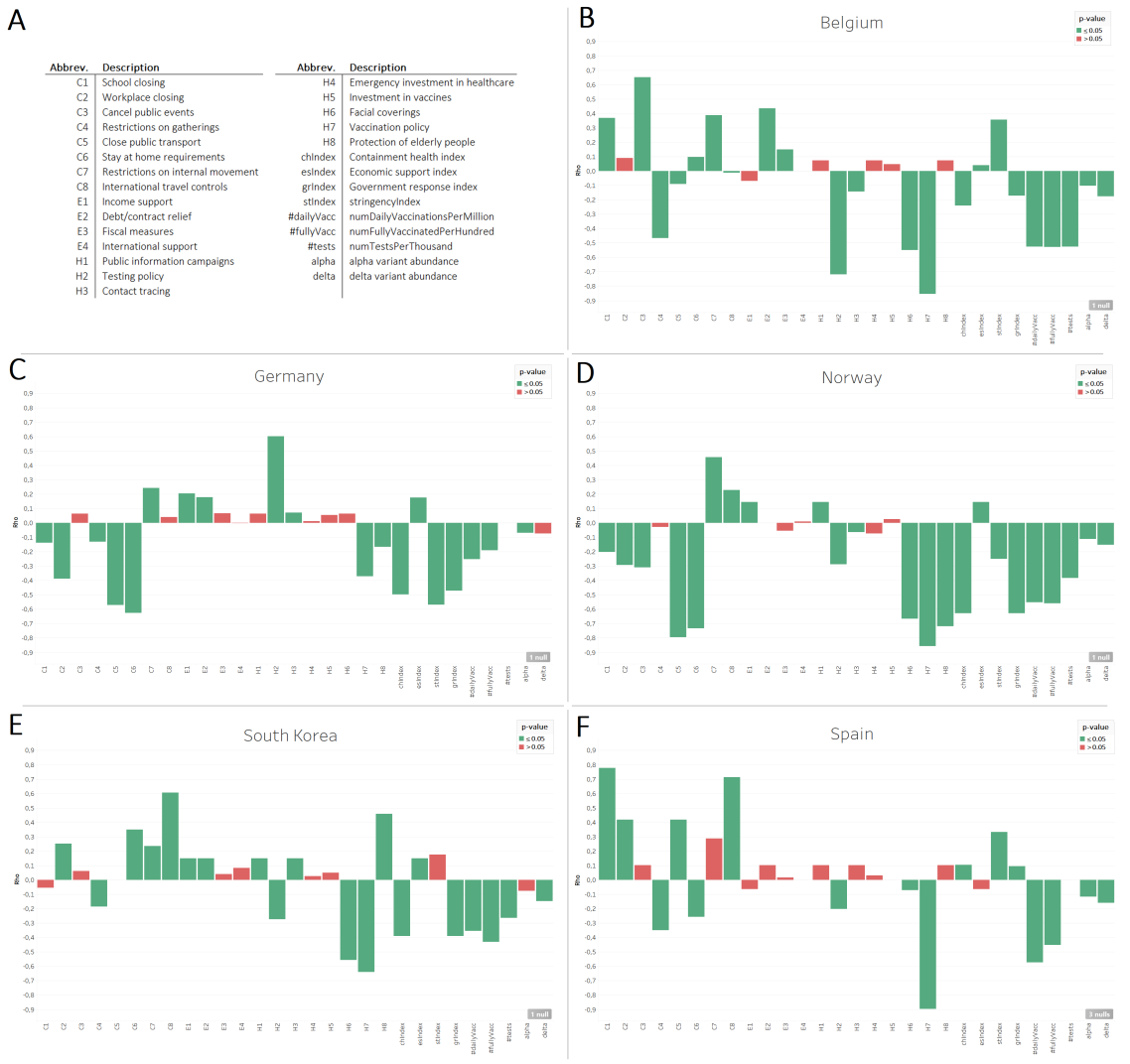

Figure 17: Results for the representative countries from Spearman analysis. A: Abbreviations of variables and policies used in the Spearman analysis. **B-F**: Rho (or ρ) values representing the relationship between each field and the mortality rate 21 days later. A positive value means that a high field value is associated with a high mortality rate; a negative value means a high field value is associated with a lower mortality rate. Bar colours indicate the confidence of these relationships: significant results (p-value ≤ 0.05) in green, and others (p-value ≥ 0.05) in red. Some countries have not reported data for all fields and are represented by null values in this figure.

present negative correlation in Spain. This could indicate differences in the implementation of the restrictions adopted.

Additionally to the correlation between policies and mortality rate, Figure 17 also presents results for the correlation between mortality rate and different metrics, like the number of vaccinated people daily and number of fully vaccinated people, as well as the number of test relative per 1000 people and the abundances of the two major COVID-19 mutations: Alpha and Delta. From this data it can be seen, that, as mentioned previously when discussing the policies regarding vaccination, for all countries there is a strong negative correlation between vaccinations conducted and mortality rate. This proves especially the effectiveness of vaccines in decreasing the severity of the disease. Furthermore, for all countries there is an overall weak negative correlation between the two variants and the mortality rate. While the data is statistically significant for these values, the magnitude of the correlation is very low. Nonetheless this could point to the two variants being potentially less deadly than the original variant of COVID-19. However, this must be viewed with

caution because, especially during the spread of the delta variant, vaccination was available to the majority of populations in the countries discussed here, which influences the data.

Finally, it is worth looking into the positively correlated policies, which could indicate ineffectiveness over the course of the pandemic. For Belgium it is the cancellation of public events (C3) and implantation of testing policies (H2) in Germany. In Norway the strongest positive correlation is between mortality rate and restrictions on internal movement (C7). In South Korea it is international travel controls (C7), while in Spain the strongest correlation is present for school closings (C1). Overall, it seems highly unlikely, that these measures had no effect on the containment of the pandemic. It could point to some co-dependencies between policies, which would imply, that the time at which a policy is implemented is crucial for its effectiveness. Due to the progression of the whole pandemic being analysed it is not possible in the scope of this analysis to gather information on the date of certain policies being implemented and the effects on the mortality.

5 Interpretation and recommendations

As mention in Section 2.2.5 of the CRISP-DM method, it is important to evaluate the results of the method used. Throughout this section, the interpretation of our results will form the basis for an implementation plan.

5.1 Interpretation

This section will see the interpretation of the final Spearman correlation analysis, which details the relationships between the various restrictions, policies, and measurements included in the analysis, and the COVID-19 mortality 21 days later. The process is illustrated in Figure 18.

A

B

Negatively correlated		Positively correlated		Varying correlation
in all five countries	in a majority of countries	in all five countries	in a majority of countries	50/50 positive and negative correlation
C4 Restrictions on gatherings H6 Facial coverings H7 Vaccination policy No. daily vaccinations No. fully vaccinated No. tests Alpha variant abundance Delta variance abundance	C5 Close public transport C6 Stay at home requirements H2 Testing policy H8 Protection of the elderly	C7 Internal movement control E1 Income support E2 Debt/contract relief E3 Fiscal measures H1 Public information campaigns	C8 International travel controls	C1 School closing C2 Workplace closing C3 Cancel public events H3 Contact tracing

Figure 18: Categorisation of policies and measures based on Spearman correlations. **A**: summary of the results from Figure 17, comparing the signs of the significant ρ per field for the five countries. **B**: classification of fields from window A into five categories, based on their correlation with mortality: strictly or majoritively negative (orange); strictly or majoritively positive (green); and varying (grey).

Firstly, Figure 18A represents a comparison of the ρ values across each field, for the five representative countries. From this, the fields' relative effectiveness in mitigating mortality can be

inferred. For instance, the policy *C4 Restrictions on gatherings* can be seen as having a negative correlation with mortality across all countries with significant results. On the other hand, *C6 Stay at home requirements* only had negative correlations with mortality in three of the five countries with significant results. Continuing these investigations for all 29 fields, a categorisation based on the distribution between negative and positive correlations emerges. These are illustrated in Figure 18B.

Eight fields have negative signs in all instances where the correlation with mortality was significant (orange box in Figure 18B). Some of these make more immediate sense than others. Firstly, it is no surprise that the restriction limiting the gathering of people is associated with a lower mortality. Preventing transmission of the virus ensures that fewer can develop severe disease that could result in death. Similarly, the use facial coverings seem to prevent deaths, also likely by preventing transmission. Thus these are the NPIs that seem the most effective in mortality mitigation, representing the main answer to the second objective of the project.

The pharmaceutical interventions are represented by the three vaccine-related fields: the vaccination policy, the number of vaccinations, and the number of fully vaccinated people. Even though they might seem to convey the same information, some specifying points can be made from them. Notably, the number of vaccinations includes both first and second vaccine doses, whereas *fully vaccinated* only counts second doses. Comparing the mean ρ for these two (see Figure 19, p. 33), it becomes apparent that *daily vaccinations* has the lower mean. This means that *daily vaccinations* has a stronger negative correlation with mortality than *fully vaccinated*, which indicates that the first vaccine dose is more effective in mitigating mortality than the second dose. However, this does not consider factors such as age or preexisting conditions, in which cases a second dose could prevent mortality to a greater extent than the first dose could in a healthy young adult, for instance. As for the third vaccination field, the H7 policy indicates the availability of vaccines, with the extreme values 0 and 5 representing no and full availability, respectively (see Table 8, p. 42). In accordance with the previous discussion, making the vaccine available to all is further emphasised as essential for mortality mitigation. Thus the third project objective is answered, indicating that vaccination is among the most effective measures in preventing mortality: more so than many NPIs, such as those discussed next.

Other measures that were strictly negatively correlated with mortality, were the number of tests, and abundance of Alpha and Delta variants. Because mortality has been calculated as deaths per case and that most COVID-19 cases do not result in death, the mortality can be reduced by simply increasing the number of infected people. This can be achieved by having more people become infected by the Alpha or Delta variants, and then receive a positive result on a test. Still, ensuring more COVID-19 cases is not the desired solution to decreasing mortality, thus the variant fields will not be further regarded.

The second category extracted from Figure 18B are fields with majoritively negative correlations with mortality. They regard public transport, stay at home requirements, protection of elderly people, and testing policy. All directly aim to mitigate viral transmission, except for the latter. Testing policy is similar to the vaccination policy in terms of representing an *availability*, rather than a restriction level. Thus the H2 field will be not be regarded further, seeing as the field *number of tests* is strictly negatively correlated with mortality.

For the remaining policies, they can also be assessed as NPIs with relevance for mortality mitigation, seeing as this has been their effect in a majority of the countries analysed. Still, a possible reason for them not being *strictly* negatively correlated with mortality can be their severity and feasibility. Demanding that inhabitants stay at home or refrain from utilising public transport are bigger interventions than requiring the use of facial covering, for example. This could lead to a government showing greater hesitancy to implement these policies given a less severe situation. Thus reserving these policies for when the transmission and mortality are high, could make them appear correlated with higher trends. This is especially true if their implementation is a result of an experienced increase in cases and deaths, rather than as a preparatory measure. Then, the transmission would already have occurred in many instances, and disease could thus develop during the period with elevated stringency. The shift of mortality by 21 days seek to combat this issue, although the effect and correctness of this shift will be discussed further in Section 5.3 (p. 35).

The strictly and majoritively positively correlated fields (green box in Figure 18B) comprise the third and fourth category of fields. They both display, as the previous groups, some less obvious categorisations. Firstly, the strictly positive fields include policies which are more relevant in a more severe transmission and mortality situation, such as *public information campaigns* and various economy-supporting policies. The internal and international travel restrictions might also be more relevant when mortality is high, for example due to the impairment these policies pose for individuals; and their limited feasibility in terms of ensuring that the restrictions are followed. Additionally, they do not prevent transmission fully, as spread can still occur within a country or region. Therefore, such policies can be envisioned as more effective in an early phase of a pandemic, or to prevent transmission to areas with no or highly controlled occurrences of cases. A final alternative for these policies, which is especially true for C8 which was negatively correlated with mortality in one country, is to evaluate the implementation strategy of these policies. Perhaps their mortality-mitigating effect would be more substantial had they been implemented earlier, for example. By reviewing the strategy, both where the intended effect was achieved and other places, these policies might be improved until they are needed anew.

The final category of policies that can be interpreted from the Spearman results, are those that had just as many positive and negative significant correlations with mortality (green box in Figure 18B, p. 30). Here, *H3 Contact tracing* will be disregarded for the same reason as the variants and testing fields. The remaining three policies have had, similarly to the majoritively positively correlated field, the intended effect in some areas. In contrast however, which factors are affecting the outcome of their implementation might be less clear due to the perfectly even distribution. They could be ineffective, implemented inappropriately, or both. A relevant observation is that they all seek to limit the gatherings, either that be in schools, at workplaces, or events. As previously seen, *C4 Restrictions on gatherings* is strictly negatively correlated with mortality. The three policies with varying correlation can only prevent gatherings in specific situations and locations, thus transmission and disease could still occur. This might point to a possible dependence between restrictions, which will be discussed in Section 5.3 (p. 35).

Overall, the interpretation of the Spearman results yield three broad classes of fields: negatively correlated (all or a majority), positively correlated (all or a majority), and varying correlation with mortality. This division will be utilised further to develop a recommended implementation plan for the WHO.

5.2 Recommended measurements and implementation plan

Among the tasks of the WHO, which were detailed in Section 1.1 (p. 1), is "*research and technical cooperation functions*". This includes dealing with global health emergencies, for which the PIP framework can be utilised. The framework describes in detail how governments should respond to the different phases of an influenza pandemic. Since there is currently no way of accurately predicting when and where a new pandemic may arise [25], no concrete timeline can be created for when specific restriction policies should be implemented to mitigate mortality. However, it is possible to define when different categories of policies should be implemented with regard to different phases of a pandemic, similarly to the PIP framework. Such a PCP framework based on COVID-19 data would be of use to all stakeholders previously mentioned. The WHO would have a clearer strategy adapted to the specific virus and disease caused by SARS-CoV-2. The advise WHO could provide would then benefit individual nations, and from there their inhabitants, businesses, and other entities.

To aid the formulation of an implementation plan describing which policies and measures should be implemented during each phase of the pandemic, each measure and policy can be placed in a prioritisation matrix according to their relative Rho-values and believed feasibility. First, they were sorted based on their mean ρ, indicating their efficiency in mitigating COVID-19 mortality. Figure 19 presents these results for the strictly negative correlated measurements.

For this group, it can be inferred that the vaccination policy has the strongest monotonically decreasing relationship with mortality, which is interpreted as the largest effect in mitigating mortality. Similar calculations of means was performed for the remaining four categories described

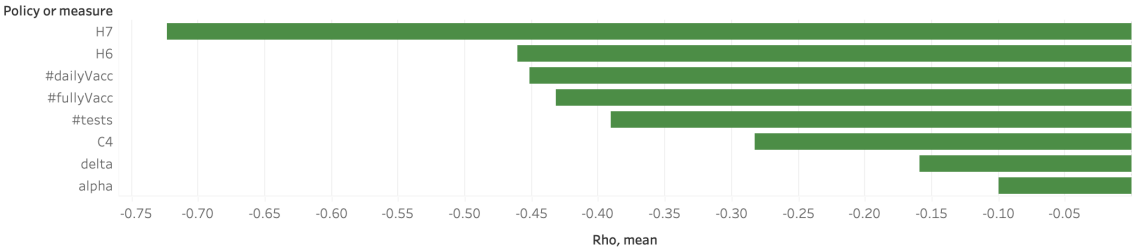

Figure 19: Mean ρ for the policies and measurements with strictly negative correlation with mortality (orange box in Figure 18). H7: Vaccination policy. H6: Facial coverings. C4: Restrictions on gatherings.

in Section 5.1 (p. 30). The means were standardised on a scale [-1, 1], with the standardised ρ yielding the y coordinate of each measurement in the prioritisation matrix.

Secondly, the relative feasibility of each measurement was determined based on their believed implementation cost, required effort, and experienced invasiveness. For instance, *stay at home requirements* is ranked low: restricting the movement of inhabitants to this extent is regarded as invasive, and the cost of hindering attendance at workplaces, schools, and social gatherings is high, both with regards to economy and well-being. On the other hand, *public information campaigns* demand relatively low implementation cost and effort, and are not regarded as invasive. The relative feasibility scores thus yielded the x coordinates in the prioritisation matrix shown in Figure 20.

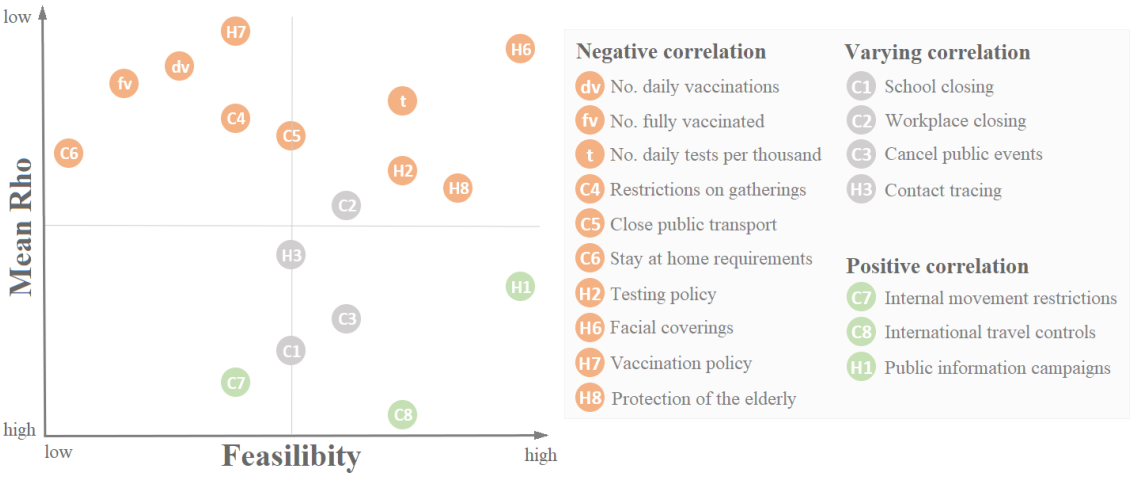

Figure 20: Prioritisation matrix plotting the feasibility of a policy against its mean ρ. A low mean ρ is associated with a larger negative relationship with COVID-19 mortality, thus the inverted y-axis. The colours correspond to the categorisation introduced in Figure 18.

From the prioritisation matrix in Figure 20, the policies and measurements can more readily be assigned their place in an implementation plan, which ultimately aims to satisfy the WHO task of responding to emerging diseases. To formulate such a framework, a simplified version of the PIP framework is utilised and combined with the results from Section 4.4 (p. 27). In the implementation plan, the pandemic phases taken into account are those presented in Table 1 (Section 1.1, p. 2): phases 1-6, post-peak period, possible new wave, and finally the post-pandemic period. The same phases utilised for Influenza viruses can be assumed for the Coronaviridae virus family, as they possess many of the same features. Examples include rapid mutation rate, viral zoonoses (i.e. transfers through animal reservoir), and seasonality.

To assign which policies and measurements should be implemented in each phase, the prioritisation matrix (Figure 20, p. 33) indicating feasibility and mean ρ were utilised. The categorisation is presented in Table 4 (p. 34).

Table 4: Assigning analysed COVID-19 measurements into policy categories for the implementation framework.

Category	Description	Measurements included
Surveillance	Situation monitoring and assessment. Initiate pandemic readiness for suspicious activity.	None
Vaccine policies	Research on vaccine production, distribution, and deployment.	H7 Vaccination policy
Local moderate policies	Increase readiness for epidemic response. Mild to moderate measurements to mitigate transmission.	H1 Public information campaigns H2 Testing policy H6 Facial coverings H8 Protection of the elderly
Local strict policies	Rapid containment of local area. Increase readiness for pandemic response.	Local moderate (H1, H2, H6, H8) C4 Restrictions on gatherings C5 Close public transport C6 Stay at home requirements C7 Internal movement restrictions H3 Contact tracing
Global moderate policies	Pandemic response. Implementation of mild to moderate measurements.	Local moderate (H1, H2, H6, H8) C8 International travel controls
Global strict policies	Imminent pandemic response.	Local strict (C4, C5, C6, C7, H1, H2, H3, H6, H7) C1 School closing C2 Workplace closing C3 Cancel public events C8 International travel controls

The first category, *Surveillance*, is not assigned any measurements from the present work as it regards a situation before human–to–human viral transmission occurs. Here, the PIP framework suggests producing, implementing, and exercising national PIP and response plans. In addition, it is advised to monitor possible origins of new or reemerging viral strains [7].

Vaccine policies is defined as a separate category due to the time required to develop vaccines. Relevant research should therefore be initiated as soon as there is evidence of human infection by a given disease. The only relevant measurement from the current analysis however, is *H7 Vaccination policy*. Its implementation ensures the establishment of a vaccine distribution prioritisation, so that vaccination can be initiated as soon they are made available.

Local moderate policies should be implemented as soon as there is evidence of pathogenic viral strains present in a community. An important policy at this stage is *H1 Public information campaigns*, in order to enlighten the public about the situation and the possible risks of infection. The moderate policies should also prompt some caution in an effort to mitigate transmission early on. These include *H2 Testing policy*, *H6 Facial coverings*, and *H8 Protection of the elderly*. From the prioritisation matrix (Figure 20, p. 33), it is clear that these measurements all have high feasibility scores, meaning they are relatively non-invasive and/or have low implementation costs. They also have negative mean ρ values, which indicates that they have been generally effective in mortality mitigation in the five studied representative countries.

If the disease situation develops further, implementation of additional and stricter local policies should be considered. If the disease progresses to sustained human–to–human transmission (phase 4 in the PIP framework), there is a higher tolerance for more costly and invasive measurements. Thus the *local strict policies* category includes measurements with lower feasibility: *C4 Restrictions on gatherings*, *C5 Close public transport*, *C6 Stay at home requirements*, *C7 Internal movement restrictions*, and *H3 Contact tracing*. In addition to the relatively low feasibility however, only the first three have a negative mean ρ. Although the effect of *C7 Internal movement restrictions* in mortality mitigation is not certain after the present analysis, it may still prevent transmission between regions. Thus is can be implemented until additional studies on its effect and implementation strategies are conducted. The final measurement included in this policy category is *H3*

Contact tracing, which should be implemented so that individuals that have been exposure to the virus are alerted.

When disease transmission is no longer contained within a community, *global moderate restrictions* may be implemented in areas that have little to no spread. These are the same as the local moderate policies: they have high feasibility and low mean ρ values, and act as preventative measures to limit further transmission. Additionally, the category contains *C8 International travel controls*. Even though an initial transmission across nation borders has occurred, continued spread of both existing and new viral strains should be limited.

Finally, the last category in the framework is *global strict policies*. It represents a full-on pandemic with prominent case numbers in multiple countries: any disease progressing to this stage requires an imminent response. Therefore, this category includes the same low-feasibility measurements as found in the *local strict policies* category, in addition to *C8 International travel controls*. Lastly, this category includes the policies that regard the closing of schools, workplaces, and public events. In the present analysis, these were found to have a varying relationships with mortality. Due to the uncertainty associated with their efficiency, as well as their relatively low feasibility, they are recommended to be reserved for this severe pandemic situation. That way there are other policies in place already, and enough time may have passed for there to have been additional studies investigating their efficiency and optimal implementation strategy.

The resulting implementation plan for PCP can be viewed in Figure 21. Along with the assigned measurements for each policy category in Table 4 (p. 34), it presents which policies should be implemented in each of the nine periods before, during, and after a pandemic. The presented implementation plan needs to be maintained and updated as new data becomes available regarding policies, restrictions and mortality rates. This is especially of importance, as soon as the world moves into the post pandemic phase, where a retrospective analysis might be easier to conduct.

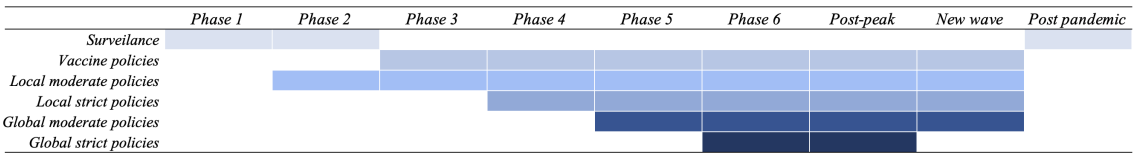

Figure 21: Gantt chart depicting the recommended implementation plan of COVID-19 policies through the pandemic: phases 1-6, post-peak, new wave, and post-pandemic stage.

5.3 Limitations and future work

As with any large project, there are limitations and possible angles to further investigate the regarded topics in future works. This subsection will highlight some limitations of the present work, with regards to choices on datasets, methods, and the model. Additionally, this subsection will provide insight into a possible extension of the project, which could see the implementation of more features that affect mortality, investigate dependencies between measurements, and the inclusion of more countries.

5.3.1 Limitations of the datasets

A well-known statement in statistics is that *"essentially, all models are wrong, but some are useful"* [26]. A key factor in determining the usefulness of a given model, is the data or premises from which it is created.

In the present work, the OxCGRT dataset important both by being a large part of the model, and also key in determining which results can be interpreted from it. A key limitation to this dataset is that it is manually curated by multiple volunteers, from countries all over the world. Even though it was found to not be significantly affected by subjective perceptions [2], there is always risk associated with having multiple curators. Furthermore, many of the the indices and policies were

based on set scales, which in many cases is not sufficiently descriptive. For instance, Australia does not have a higher stringency than the global mean, which does not seem representative considering how long the country has had strict regulations in place. This can be seen in Figure 2 (p. 13). A possible reason for this observation is the small differences in policies that are not tracked in the OxCGRT dataset. Likewise, Sweden did not have a significantly lower stringency compared to other countries, despite the country's relatively relaxed restriction response.

The dataset on virus variant abundances was only updated every two weeks, with the earliest entry being as late as May 2020. Additionally, it only covers 95 countries. This made the variants dataset quite sparse, and thus difficult to use in combination with the other datasets with higher coverage. Thus variant data has not been emphasised in the present project. Acquiring more data on variants would remedy this limitation however, and could enable additional interesting analyses, such as the early project idea of predicting variant abundances.

Even though data entries reporting mortality rates of 100% were considered outliers in visualisations such as Figure 5 (Section 4.2.1, p. 19), they illustrate the limitation of relying on mortality as a measurement of the number of deaths divided by the number of COVID-19 infected. For the earliest cases and reports of deaths, large deviations occurred between countries because of a lack of reported testing data. This was most likely a result of a reduced testing capacity, as countries with a low testing capacity often had a higher mortality. Having an infection fatality rate instead of mortality would therefore have been beneficial. However, no accurate data on the infection fatality rate for the beginning phases of the pandemic was available when the project was initiated, thus the team settled for an estimate.

5.3.2 Limitations in method and model

A Spearman rank-order correlation analysis was used to compare the relationship between mortality and different restrictions and policies. The analysis is, as previously mentioned, appropriate for data that is not normally distributed, and it did produce interesting results for the policies that were negatively correlated with mortality. However, for the policies that had a positive or varying correlation in the representative countries, a clear interpretation of their effect and value is more difficult to provide. Therefore, using only the Spearman correlation as the modelling foundation for the PCP framework is not ideal. However, through the iterations of CRISP-DM, implementing another model was decided against due to the time constraints of the project.

There are also certain limitations to the mortality shift. As the approach for generating the shift was justified using short time frames, the effects on long time frames are not known. In rudimentary investigations, the magnitude of the negative correlation between shifted mortality and original data was lessened for all countries. At the same time, the correlation between original data and the stringency index is lesser for long time frames, than for short. This makes trends more obscure, due to large variations within the stringency measurements. Still, the benefits seem to outweigh these limitations, further supporting the choice to shift the mortality data.

Both the choice of groups and representative countries were based on biased initial hypotheses. An important factor is the broadcast of the pandemic during its initial phases, which the group members experienced through local news and social media. Therefore, the groups defined in the project might be more diverse than if they had been based on other criteria. For instance, the categorisation could primarily be based on age demographics and political, social and healthcare systems. In addition, the grouped countries of this project are based around Europe, with the exception of the SARS group. This in turn might limit the application of the PCP framework to the countries within Europe, unless more countries are added to the model.

5.3.3 Future work

Firstly, future expansions of the project should address the mentioned limitations and present solutions as to how they can be improved. With regard to the OxCGRT dataset, future work could include creating a restriction and policy dataset which records nuances on a larger scale.

This could be done by for instance by increasing the diversity of possible values for each policy. For instance, *C5 Close public transport* currently only allows three possible values: no measure; recommended closing, or required closing. Expanding the value range to five could for instance see the inclusion of partial closing of selected routes or transportation mediums.

Future tasks could also include additional restrictions and policies, as well other factors that may affect mortality. It could be interesting to look into fields concerning geographical movement that affects spread of the disease, as well as utlising the demographics data to a greater extent. Examples include population mean age and density.

Furthermore, investigating potential dependencies between policies could yield insights into which policies are actually affecting mortality. One policy could for instance always increase or decrease in cohesion with another, but not directly influence the mortality itself. The current analysis cannot identify such dependencies between policies, but would rather yield results indicating that both are effective in mortality mitigation. By taking this into account in future work, insights concerning the preferred order in which policies are introduced could be gained. Another dependency that may be interesting to investigate is between variables in the demographic data and COVID-19 mortality.

The current analysis only includes five groups of countries, with one representative chosen for each group. Most of these consist of European countries, with the only exception being the SARS group. Future expansions should look into developing a model with groups more representative for the world population, or utilise an analysis of all countries in an automated manner. Thus any limitations and possible issues of selecting subsets of countries would be avoided.

6 Conclusion

The present work sought to investigate the relationship between COVID-19 mortality and governmental imposed restrictions and policies, and to utilise these results to counsel the WHO regarding a PCP framework. The conducted analyses indicated that there was a correlation between stringency and mortality. Some policies were found to have monotonically decreasing relationships with mortality, indicating that they have contributed in mitigating COVID-19 mortality. Examples include policies on vaccinations and testing, the use of facial coverings, and restrictions on gatherings. These and other findings were used to formulate a PCP framework, which could aid WHO in recommending restrictions to government in the event of another re-emerging COVID-19 pandemic. In light of the limitations the datasets, model, and methods, possible extensions or continuations of the present work should focus on improving the model and implementing more countries to increase its global relevancy.

Bibliography

[1] The World Health Organization. Who coronavirus (covid-19) dashboard, 2021. [Internet]. WHO. Updated 05.10.21. Retrieved 05.10.21 from https://covid19.who.int/.

[2] M. J. Mækelæ, N. Reggev, N. Dutra, R. M. Tamayo, R. A. Silva-Sobrinho, K. Klevjer, and G. Pfuhl. Perceived efficacy of covid-19 restrictions, reactions and their impact on mental health during the early phase of the outbreak in six countries. *Royal Society open science*, 7 (8):200644, 2020.

[3] Folkehelseinstituttet. Koronavaksinasjon - statistikk, 2021. [Internet]. FHI. Updated 05.10.21. Retrieved 05.10.21 from https://www.fhi.no/sv/vaksine/koronavaksinasjonsprogrammet/koronavaksinasjonsstatistikk/.

[4] T. R. Frieden, M. Buissonnière, and A. McClelland. The world must prepare now for the next pandemic. *BMJ Global Health*, 6(3):e005184, 2021.

[5] The World Health Organisation. What we do, 2021. [Internet]. WHO. Retrieved 27.10.21 from https://www.who.int/about/what-we-do.

[6] D. Ruger, JP. Yach. The global role of the world health organization. *Global health governance: the scholary journal for the new health security paradigm*, 2(2):1–11, 2009.

[7] World Health Organization. Pandemic influenza preparedness and response, 2009. [Internet] Retrieved 18.11.21 from https://apps.who.int/iris/bitstream/handle/10665/44123/9789241547680_eng.pdf?sequence=1&isAllowed=y.

[8] V. C. C. Cheng, S. K. P. Lau, P. C. Y. Woo, and K. Y. Yuen. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clinical Microbiology Reviews*, 20(4):660–694, October 2007.

[9] Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, and Phillips T. A global panel database of pandemic policies (oxford covid-19 government response tracker). *Nature Human Behaviour*, 5:529–538, 2021.

[10] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and W. Rüdiger. Crisp-dm 1.0: Step-by-step data mining guide, 2000. [Internet] Retrieved 10.11.21 from https://www.the-modeling-agency.com/crisp-dm.pdf.

[11] G. A. Hindle and R. Vidgen. Developing a business analytics methodology: A case study in the foodbank sector. *European Journal of Operational Research*, 268(3):836–851, 2018.

[12] A. T. Ferreira, C. Fernandes, J. Vieira, and F. Portela. Pervasive intelligent models to predict the outcome of covid-19 patients. *Future Internet*, 13(4):102, 2021.

[13] R. Wirth and J. Hipp. Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, volume 1. Springer-Verlag London, UK.

[14] S. Shen. 7 steps to ensure and sustain data quality, 2019. [Internet]. Towards Data Science. Retrieved 18.11.21 from https://towardsdatascience.com/7-steps-to-ensure-and-sustain-data-quality-3c0040591366.

[15] H. Lu, C. W. Stratton, and Y. Tang. Outbreak of pneumonia of unknown etiology in wuhan, china: The mystery and the miracle. *Journal of Medical Virology*, 92(4):401–402, 2020.

[16] Worldometers. Countries in the world by population, 2021. [Internet]. Retrieved 28.09.21 from https://www.worldometers.info/world-population/population-by-country/.

[17] S. Boccia, W. Ricciardi, and J. P. A. Ioannidis. What other countries can learn from italy during the covid-19 pandemic. *JAMA Internal Medicine*, 180(7):927, 2020.

[18] Preda G. Covid-19 world testing progress, 2021. [Internet]. Kaggle. Updated 28.09.21. Retrieved 28.09.21 from https://www.kaggle.com/gpreda/covid19-world-testing-progress.

[19] Preda G. Covid-19 variants worldwide evolution, 2021. [Internet]. Kaggle. Updated 16.09.21. Retrieved 28.09.21 from https://www.kaggle.com/gpreda/covid19-variants.

[20] Sahni S. Covid-19 world vaccination progress, 2021. [Internet]. Kaggle. Updated 08.08.21. Retrieved 28.09.21 from https://www.kaggle.com/sagarsahni3/covid19-world-vaccination-progress.

[21] Tableau Software LLC. Tableau prep builder, 2021. [Software]. Version 2021.3.1. Retrieved 27.09.21 from https://www.tableau.com/products/prep.

[22] D. Baud, X. Qi, K. Nielsen-Saines, D. Musso, L. Pomar, and G. Favre. Real estimates of mortality following covid-19 infection. *The Lancet infectious diseases*, 20(7):773, 2020.

[23] Tableau Software LLC. Tableau desktop public edition, 2021. [Software]. Version 2021.3.1. Retrieved 27.09.21 from https://public.tableau.com/en-us/s/download.

[24] R. A. Armstrong. When to use the bonferroni correction. *Ophthalmic and Physiological Optics*, 34(5):502–508, 2014.

[25] M. Wille, J. L. Geoghegan, and E. C. Holmes. How accurately can we assess zoonotic risk? *PLOS Biology*, 19(4):e3001135, 2021.

[26] G. EP. Box and N. R. Draper. *Empirical model-building and response surfaces.*, page 424. John Wiley & Sons, 1987.

[27] P. Schober, C. Boer, and L. A. Schwarte. Correlation coefficients: Appropriate use and interpretation. *Anesthesia Analgesia*, 126(5):1763–1768, 2018.

[28] S. S. Shapiro and M. B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

[29] F. C. Lam and M. T. Longnecker. A modified wilcoxon rank sum test for paired data. *Biometrika*, 70(2):510–513, 1983.

Appendix

A Overview of datasets

This appendix provides overviews of all data fields available in the raw datasets utilised in the project. Even though only a subset of these were actively used, all have been included here to highlight the sets' versatility for further studies, and the presence of specific sources for each data entry.

A.1 COVID-19 Testing Progress

The full dataset [18] contains the fields described in Table 5. Utilised categories for the project are country and date (for integration purposes), and the number of daily tests per thousand inhabitants, see Section 3.1.1 (p. 11).

Table 5: All fields in the dataset *COVID-19 World Testing Progress*.

Original field	Renamed field	Description
Entity	testUnit	Originally the name and the testing unit utilised for that country. Split into two separate fields; country and testing unit. Test units are either *tests performed*, *people tested*, *samples tested*, or *units unclear*.
Source Label	testSource	Source name for the data entry
Notes	testNote	Optional field, may include remarks to the data entry
Daily change in cumulative total	#dailyTests	Increase in test unit per day per country
Cumulative total	#testsTotal	Total count of test unit until this date per country
Cumulative total per thousands	#testsTotalPer Thousand	total count of test unit until this date per country, per thousand inhabitants
Daily change in cumulative total per thousands	#dailyTestsPer Thousand	Increase in test unit per day per country, per thousand inhabitants
7-day smoothed daily change	#tests7daySmooth	Daily change in test unit per day per country, average over the last seven days
7-day smoothed daily change	#tests7daySmooth PerThousad	Daily change in test unit per day per country per thousand inhabitants, average over the last seven days
Short-term positive rate	positiveRate	Rate of positive results from tests per day per country, average over a short period
Short-term tests per case	#testsPerConfirmed Case	Count of test unit per confirmed COVID case per day per country, averaged over a short period

A.2 COVID-19 Variants Worldwide Evolution

This dataset [19] details the relative abundances of SARS-COV-2 variants, whose fields are all described in Table 6. The two fields utilised in the project in addition to date and country, were the percent abundances of the Alpha and Delta variants, see Section 3.1.2 (p. 12).

Table 6: Fields from *COVID-19 Variants Worldwide Evolution* utilised in *masterDatasetName*

Original field	Renamed field	Description
num_sequences	#sequencesTotal	Total number of sequences per date per country
perc_sequences	*22 transponsed fields* %*variant name*	Relative abundance of variants *alpha*, *B.1.1.277*, *B.1.1.302*, *B.1.1.519*, *B.1.160*, *B.1.177*, *B.1.221*, *B.1.258*, *B.1.367*, *B.1.620*, *beta*, *delta*, *epsilon*, *eta*, *gamma*, *iota*, *kappa*, *lambda*, *mu*, *S:677H.Robin1*, *S:677P.Pelican*, and *other*
perc_sequences	*1 transponsed field* %nonWHO	How many percent of sequenced variants are known to WHO, per date per country

A.3 COVID-19 World Vaccination Progress

An overview of the complete dataset [20] is given in Table 7. Included in *masterDatasetName* are the fields on the number of daily vaccinations and fully vaccinated inhabitants, see Section 3.1.3 (p. 12).

Table 7: Fields from *COVID-19 World Vaccination Progress*, data sheet *Country Vaccinations* utilised in *masterDatasetName*

Original field	Renamed field	Description
Total number of vaccinations	#vaccinations	Cumulative total of vaccinations administered until that day per country
Total number of people vaccinated	#peopleVaccinated	Cumulative total of people vaccinated until that day per country. In the case of all inhabitants receiving two doses, this value will be double the population number
Total number of people fully vaccinated	#fullyVaccinated	Cumulative total of people fully vaccinated according to the country's vaccination scheme (typically two each)
Daily vaccinations (raw)	#dailyVaccinations Raw	Number of vaccinations per day per country
Daily vaccinations	#dailyVaccinations	Number of vaccinations per day per country
Total vaccinations per hundred	#vaccinationsPer Hundred	Percent ratio of vaccination number and population, until that day per country
Total number of people vaccinated per hundred	#peopleVaccinated PerHundred	Percent ratio of people vaccinated and population, until that day per country
Total number of people fully vaccinated per hundred	#fullyVaccinated PerHundred	Percent ratio of people fully vaccinated and population, until that day per country
Daily vaccinations per million	#dailyVaccinations PerMillion	Ratio of daily vaccination number and population per day per country
Vaccines	allVaccinesUsed	All vaccine brands used per day per country
Source name	vaccSource	Source name for the data entry
Source website	vaccURL	Source URL for the data entry

A.4 COVID-19 Government Response Tracker

Table 8 details the 20 indicators from the Oxford COVID-19 Government Response Tracker dataset, as well as the meaning of their possible values and flags. All were included in the analysis, although some were emphasised more than others. In addition to these indicators, the dataset contains indexes, confirmed deaths, and confirmed cases, which were also utilised in the project, see Section 3.1.4 (p. 12).

Table 8: Descriptions of indicators of *COVID-19 Government Response Tracker* included in *masterDatasetName*.

CONTAINMENT AND CLOSURE POLICIES				
	C1 Schools closing	C2 Workspaces closing	C3 Cancel public events	C4 Gathering restrictions
0	No measures	No measures	No measures	No measures
1	Recommended	Recommended	Recommended	If over 1000 people
2	Partial closing	Partial closing	Required cancelling	If 101-1000 people
3	Full closing	Full closing	-	If 11-100 people
4	-	-	-	If 10 or less people
Flags				
0	Local measure	Local measure	Local measure	Local measure
1	National measure	National measure	National measure	National measure
	C5 Close public transport	C6 Stay at home requirements	C7 Restrictions on internal movement	C8 International travel controls
0	No measures	No measures	No measures	No restrictions
1	Recommended	Recommended	Recommended	Screening arrivals
2	Required	Required	Required	Quarantines
3	-	-	-	Ban some arrivals
4	-	-	-	Ban all arrivals
Flags				
0	Local measure	Local measure	Local measure	-
1	National measure	National measure	National measure	-
ECONOMIC POLICIES				
	E1 Income support	E2 Debt/contract relief	E3 Fiscal measures	E4 International support
0	No support	No relief	-	-
1	$<$ 50% of lost salary	Narrow, contract-specific	-	-
2	$\geq$ 50% of lost salary	Broad relief	-	-
$	-	-	USD spent on economic stimulus	USD given in aid to other countries
Flags				
0	Formal/informal sector only	-	-	-
1	All workers	-	-	-
HEALTH SYSTEM POLICIES				
	H1 Public information campaigns	H2 Testing policy	H3 Contact tracing	H4 Emergency investment in healthcare
0	No campaign	No policy	No tracing	-
1	Officials urge caution	If with symptoms and fulfils special criteria	Limited	-
2	Coordinated campaign	If with symptoms	For all cases	-
3	-	Open public testing	-	-
$	-	-	-	USD in short-term spending
Flags				
0	Local measure	-	-	-
1	National measure	-	-	-
	H5 Investment in vaccines	H6 Facial coverings	H7 Vaccination policy	H8 Protection of elderly people
0	-	No policy	No availability	No measures
1	-	Recommended	For one group: key workers, clinically vulnerable, elderly	Recommended isolation, hygiene, visitor restriction
2	-	Some public areas when others present	For two groups	Narrow restrictions
3	-	All public areas when others present	For all three groups	Extensive restrictions
4	-	Whenever outside	For all three groups and some other groups	-
5	-	-	Available to all	-
$	USD spent on vaccine development	-	-	-
Flags				
0	-	Local measure	Full/mostly individual cost	Local measure
1	-	National measure	No/minimal individual cost	National measure

B Statistical analyses: Theory

In this section, an overview of the theory will be provided for the different statistical tests used in this project.

B.1 The Spearman's rank correlation

The Spearman's rank correlation measures the strength and direction of assoctiation between two ranked variables. The output of a Spearman's rank correlation is a Spearman's rank correlation coefficient, called ρ (rho), which varies from +1 to -1 and describes the relationship between two variables using a monotonic function. It is calculated as shown in Equation 1, for two samples of size n with scores X_i, Y_i where ρ is the Spearman correlation coefficient, σ denotes the standard deviation and cov the covariance.

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \hat{x})^2 \sum_i (y_i - \hat{y})^2)}} \tag{1}$$

The value of ρ determines the monotonic relationship between the two variable, meaning whether they increase or decrease in cohesion. The definition of different relationships for varying ρ's can be seen in Table 9. The standard way of denoting Spearman's rank correlation results is ($r(df) = \rho$, p-value), where df is the degree of freedom and ρ is the rho value [27].

Table 9: Description of relationship between two data-sets for different values of ρ

ρ	Description of relationship
1	Perfectly monotonically increasing
0.8	Strong monotonically increasing
0.2	Weak monotonically increasing
0	Non-monotonic relation
-0.2	Weak monotonically decreasing
-0.8	Strong monotonically decreasing relationship
-1	Perfectly monotonically decreasing

B.2 Shapiro-Wilk test of normality

The Shapiro-Wilk test of normality is a statistical test which investigates whether or not a sample came from a normally distributed population [28]. It does so by superimposing a normal curve over the observed sample. It then computes a similarity percentage between the normal curve and the observed sample, generating a similarity percentage. If the p-value of the Shapiro-Wilk Test is $>$ 0.05, the distribution of the sample is not significantly different from that of a normally distributed sample. However, if the p-value < 0.05, one can say that the sample is significantly different from that of a normally distributed sample of the same size [28].

Therefore the null hypothesis (H_0) is that the sample is normally distributed, whilst the alternative hypothesis (H_1) is that the sample is not normally distributed. For more information behind the calculation of the Shapiro-Wilks test, see Shapiro and Wilk [28].

B.3 Wilcoxon rank sum test

The Wilcoxon rank sum test is a non-parametric test used to compare the data of two samples. Of interest, is the "non-parametric" notation, which denotes that the test does not require any specified distribution for the samples. Therefore, a Wilcoxon rank sum test can be used on samples that are not normally distributed, in contrast to the alternative T-test which assumes normally distributed data. The T-test is therefore favoured for normally distributed data. There are two

variants of the test: one-sampled and paired sampled. The one-sample test investigates whether the data comes from population that is distributes symmetrically around a given median. Most practical applications of the Wilcoxon analysis, however, use paired samples where the two samples are tested for whether or not they are interchangeable [29].

The null hypothesis (H_0) for this test is that the medians of two samples are equal. Meanwhile, the alternative hypothesis (H_1) indicates a set of samples that do not have equal medians.

The output of a Wilcoxon analysis is a p-value. A p-value < 0.05 indicates that the null hypothesis can be rejected meaning that the two samples do not have an equal mean and are interchangable. For a p-value > 0.05, the null hypothesis cannot be rejected and the samples are [29]. For more information on the paired Wilcoxon analysis, see Lam and Longnecker [29].

C GitHub description

This section will describe GitHub where all raw data and the script for each analysis is provided. The GitHub can be found at https://github.com/SaraAsche/TDT4259-Group-25. It contains several files in the format R-markdown (Rmd), Jupyter Notebook(ipynb), excel(xlsx), and comma-separated values (csv).

Information on the approach to run Wilcoxon, Shapiro-Wilk and Spearman correlation can be found in the R-markdown file "Data Science.Rmd". This file further outputs excel files for the Spearman correlation matrix generated from the policy and restrictions fields discussed in the method section. For each country, there are two matrix outputs from the Spearman analysis. One denotes the p-value related to each output whereas the second one details the rho-value.

For information on the 21-day shift in mortality, the Jupyter file "TEST_Countries.ipynb" describes the analysis done to justify the shift in mortality.