Fraunhofer
PORTUGAL

RESEARCH CENTER FOR ASSISTIVE INFORMATION AND COMMUNICATION SOLUTIONS

FAIRisk

Improving risk estimation with open resources

EUROPEAN OPEN SCIENCE CLOUD

EUROPEAN OPEN SCIENCE CLOUD

This work was funded by EOSCsecretariat.eu. EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-05-2018-2019, Grant Agreement number 831644.

RESEARCH CENTER FOR ASSISTIVE INFORMATION AND COMMUNICATION SOLUTIONS

FAIRisk

Improving risk estimation with open resources

Authors

Diana Gomes (diana.gomes@fraunhofer.pt)

Catarina Pires (catarina.pires@fraunhofer.pt)

David Ribeiro (david.ribeiro@fraunhofer.pt)

Duarte Folgado (duarte.folgado@fraunhofer.pt)

Ricardo Santos (ricardo.santos@fraunhofer.pt)

Telmo Barbosa (telmo.barbosa@fraunhofer.pt)

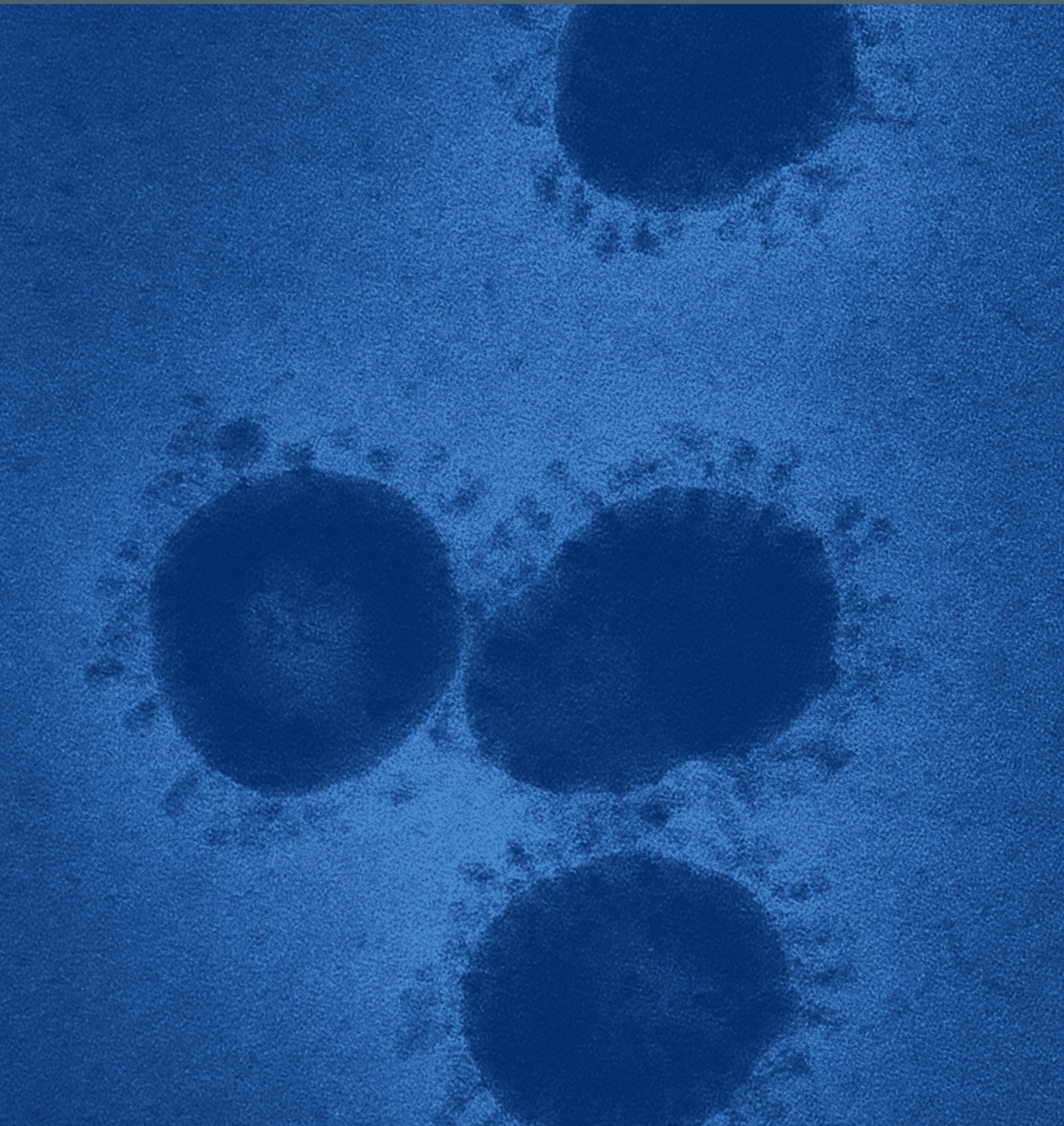
Affiliation

Fraunhofer Portugal AICOS

Fraunhofer
PORTUGAL

Version

2.0



Abstract

As a society, we know that we must be prepared for any hazards. Some of them are very well-known, like earthquakes or even civil conflicts; but some others are rather unexpected or, at least, very rare. The COVID-19 pandemic has brought the discussion of preparedness and response to crisis to the top charts. But can we use what we observed after the break of this pandemic to increase our knowledge about worldwide risk modelling?

This work used open-source globally available data from several sources, some of which are frequently used to compute risk ranks and estimate countries' preparedness for crisis, with post-COVID-19 data to create preventive insights and suggest improvements to current risk modelling strategies. Excess mortality was used as an objective measure of the overall impact of the disease worldwide, in terms of loss of lives.

Our preliminary analysis supports that there is room for improvement of current risk ranking scales, country-level attributes' data withhold hidden relations with potential to assist conceptual risk modelling expertise, and using a single model to forecast country-level excess mortality is likely to be feasible, despite being extremely limited by the amount of available data. Moreover, this analysis showcases the potential of the combined data and further steps of future work are discussed to encourage the reusability of interoperable multimodal data and adequate open-science policies.

Table of Contents

1. “Always Be Prepared”

The motto was given by Baden-Powell, but the struggle is common. The appearance of a worldwide pandemic has brought the discussion of countries’ awareness, preparedness, and response capacity to crisis to the public domain and scrutiny. COVID-19 has reminded the world of the importance of anticipating all types of hazards and devise adequate responses for the follow-up crisis. The role of open data and science has also become more evident, along with initiatives to release such data and increase its interoperability. It is in this context that the following analysis was proposed.

Over the following pages, a discussion about **how open data can contribute towards improving risk estimation worldwide** will be promoted over several different perspectives. The first step consists of analysing whether risk models currently openly available were able to foresse the COVID-19’s impact in loss of lives – the most devastating impact measure - across countries. Then, an unsupervised modeling approach using open-source data from several different sources will add value to the discussion of whether hazard-agnostic country profiling can disclose relevant hidden patterns that can assist the conceptual work of experts. Finally, we propose an excess mortality forecasting approach that considers country-level indicators, mobility and COVID-19-specific data for its predictions.

The document is organized into 5 sections: Section 1 provides the context and motivation of the disclosed work; sections 2 to 4 address our main research questions and the outcomes of our analysis; and section 5 concludes the document by offering an overall discussion of its ideas and proposing further directions for future work.

1.1. Were countries prepared for a widespread pandemic crisis?

The COVID-19 pandemic evolution showed that the majority of countries were not prepared for a widespread pandemic. In 2011, the International Health Regulations Review committee declared that *“the world is ill-prepared to respond to a severe influenza pandemic or to any similarly global, sustained and threatening public-health emergency”* and, by the beginning of 2019, less than 1 in 3 European countries had revised their pandemic plans since the 2009 swine flu pandemic, despite recommendations from the World Health Organization (WHO).

But can we even measure the level of preparedness of each country? The Global Health Security Index, for example, provides a measure of health security around the world and aims to provide an estimate of the level of countries’ preparedness for a pandemic. According to their 2019 index, the United States of America was the most well-prepared country, with a score of 83.5 out of 100, followed by the United Kingdom, Netherlands, Australia, and Canada. Although some countries obtained a high score, this index showed that the healthcare systems around the world are weak, with an average score of 40.2, and that no country was fully prepared for a pandemic.

1.2. About risk scales, indexes, and tools

Epidemic management is a highly complex topic which addresses the actions related to anticipating, preparing for, preventing, detecting, responding, and controlling epidemics to minimize their health, social and economic effects. Management of such a volatile and dynamic crisis requires adequate tools to support decision-making.

An epidemic risk index (ERI) is a mathematical tool that helps strategic decision-making and prioritizes capacity-building activities for national epidemic prevention, preparedness, and response. We present a high-level overview of some examples of open initiatives on this matter. Some properties of these initiatives are also combined and presented in Table 1 comparison.

INFORM Risk[1]

The INFORM model (Marin-Ferrer et al., 2017) developed by the Joint Research Centre of European Commission (JRC), envisages three dimensions of risk: Hazards & Exposure, Vulnerability, and Lack of coping capacity. It is split into different levels to provide a quick overview of the underlying factors leading to humanitarian risk. It builds up the picture of risk by more than 50 core indicators.

Global Health Security Index[2]

The Global Health Security (GHS) Index provides benchmarking of health security and related capabilities across the 195 countries. The GHS Index relies on data that a country has published on its own or has reported to or been reported by an international entity. The score is calculated based on data of 34 indicators grouped across six categories: prevention; detection and reporting; rapid response; health system; compliance with international norms; and risk environment.

ReadyScore Map[3]

The ReadyScore relies on data from the Joint External Evaluation (JEE) from the World Health Organization (WHO). The score is the average of the JEE technical area scores, which are subgrouped into four categories: prevention, detection, response, and other.

Facebook Risk Score[4]

Effective responses to pandemics require identifying small clusters of at-risk populations to target public health resources quickly. The Facebook Data for Good initiative released several datasets on demographic and mobility data. These datasets can be used to determine updated risk scores worldwide, considering essential risk factors, such as the percentage of the elderly population and mobility patterns evaluating mandatory curfews' compliance.

1 https://drmkc.jrc.ec.europa.eu/inform-index/INFORM-Risk
2 https://www.ghsindex.org/
3 https://preventepidemics.org/map/
4 https://dataforgood.fb.com/docs/tutorial-identification-of-at-risk-populations/

Table 1 Comparison between open initiatives related with global epidemic risk estimation.

	# Countries	Target estimation	Indicators/ Outcome	Available before COVID-19	Specific for COVID-19
INFORM Epidemic Risk	**195**	**Risk**	**Static**	**Yes**	**No**
INFORM COVID-19 Risk	**195**	**Risk**	**Static**	**No**	**Yes**
GHS Index	**195**	**Security**	**Static**	**Yes**	**No**
ReadyScore Map	**100**	**Preparedness**	**Static**	**N.A.**	**No**
Facebook Weighted Risk Score	**152**	**Risk**	**Dynamic**	**No**	**Yes**

The number of countries considered by each initiative is not consistent. This fact is mostly related to the lack of available data for the indicators considered by each method. The ReadyScore Map conveys information for the smallest number of countries, focusing mainly on developing areas. Additionally, estimations do not always foresee the same target. While some scales aim to estimate risk, the others are designed to predict its opposite, i.e., security or preparedness. The Facebook Risk Score also differs from the remaining approaches in another aspect. It was created following the COVID-19 pandemic using daily population density and mobility estimations. As such, this approach is not directly comparable with the remaining scales, which rely on self-reported, measured, or qualitative static indicators to provide a tool to predict country preparedness and response to crisis rather than model dynamic indicators specifically associated with COVID-19 spread and impact.

These examples demonstrate the efforts being carried out by the community to offer risk scales, indexes, and tools for epidemic risk management. However, reaching a representative score that summarizes such a multitude of factors is not an easy task. For instance, JEE focus on public health competencies in great depth but does not fully address the broader range of non-health system factors, including institutional, financial, and infrastructural capacities, which are also fundamental building blocks for an effective response to infectious. Additionally, some indexes consist of self-reported data, which raises the potential for bias and inaccurate reporting.

1.3. Excess mortality as a measure of pandemic impact

Excess mortality is defined as the number of additional deaths from all causes during a crisis when compared to the expected number of deaths during normal conditions (Ritchie et al., 2020). In the case of the COVID-19 pandemic, excess mortality measures the additional number of deaths from all causes over 2020 and 2021 periods by comparing the real number of deaths in those periods with the expected number of deaths, estimated from analogous periods of past years. There are several ways to estimate the number of expected deaths – the baseline. One of the most common methods is based on the average of deaths per period over the 5 years prior to the pandemic, i.e., from 2015 to 2019. These excess deaths can be understood as the deaths caused both directly and indirectly by COVID-19 (Beaney et al., 2020).

Excess mortality can also be standardized by population to allow the computation of the P-score, which facilitates comparisons between countries (Aron & Muellbauer, 2020; Ritchie et al., 2020). The P-score is the absolute number of excess deaths normalized by the baseline. Therefore, it indicates the percentage of deaths that are above normal deaths. For example, a P-score of 100% for a given week in 2020 indicates that the number of deaths on that week was two times higher than expected based on the previous years used to estimate the baseline.

The number of COVID-19 deaths is not enough to estimate the real impact of the pandemic. For instance, some vulnerable individuals who died from COVID-19 could have died from alternate causes. That is, there can be a short-term increase in mortality due to COVID-19 that results in a reduction of mortality from other causes over time, also known as mortality displacement (Beaney et al., 2020). Besides that, not all deaths of COVID-19 are accounted for - some can be assigned to other causes if COVID-19 was not diagnosed or mentioned on the death certificate. On the other hand, the total number of deaths caused by the pandemic might be even higher than the COVID-19 deaths, for example, due to the extra burden of hospitals with COVID-19 patients. This might have delayed the treatment of other health conditions or discouraged people from seeking medical treatment (Aron & Muellbauer, 2020). However, the pandemic might have lowered the deaths of some causes, such as the flu or road accidents, due to mobility restrictions (Ritchie et al., 2020).

The numbers of COVID-19 deaths also do not allow for an accurate comparison between countries because they are affected by differences in reporting and testing (Beaney et al., 2020; Ghislandi et al., 2020). For instance, some countries only report COVID-19 deaths that happen in hospitals and not at home (Krelle et al., 2020). Besides that, some countries only report the deaths of confirmed COVID-19 cases, excluding untested individuals.

Since excess mortality accounts for all causes of death and the P-score measure can provide an accurate country comparison, it can provide relevant information towards assessing and comparing pandemic's impact across countries.

1.4. How can open data help?

The open share of data during a pandemic is very significant. In addition to allowing the research community to receive expedited access to relevant data, it promotes transparency and reliability. It also fosters interdisciplinary collaboration on understanding the pandemics' dynamics and facilitates coordinated and timely responses at an international level.

As the volume, complexity, and creation of accessible data increases, the need for guidelines to adequately release and maintain a good quality of data becomes even more critical. In 2016 the *'FAIR Guiding Principles for scientific data management and stewardship'* were proposed to provide guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets.

A dataset that unifies several dimensions would be an added value, allowing to express risk indexes based on a comprehensive description. Furthermore, modeling approaches should also take into consideration the principles of fairness and transparency. That was the goal behind the proposition of the *FAIRisk* repository[5]. This initiative addresses data interoperability challenges of fetching and combining several openly available third-party sources of multimodal data, so these can be coherently used in a unified data model. This model was designed bearing FAIR principles and EU's open data guidelines in mind to promote an adequate, well-documented and simplified use of the combined data.

5 https://github.com/fraunhoferportugal/fairisk

These resources were used to fetch the data from the sources and generate all data selection and transformations required to reproduce the results reported in this document. We used a static version of the data to conduct the experiments (local JSON file created after data fetch from all sources documented in *FAIRisk's* repository on March 23rd, 2021).


Total Deaths
4.720
68.324
RUSSIA
3.056 deaths
Hubei China
827 deaths
Italy
429 deaths
Iran
66 deaths
Korea, South
55 deaths
Spain
48 deaths
France France
31 deaths
Washington US
22 deaths
Henan China
16 deaths
Japan
13 deaths
Heilongjiang China
Hubei China
Iran
Guangdong China
Henan China
Zhejiang China
Italy
Hunan China
Anhui China
Jiangxi China
Shandong China

2. Should Conceptually Modelled Static Risk Scales Be Improved?

The risk scales, indexes and tools introduced and compared in Table 1 were created to assist both the detection of dangerous soft spots and the decision-making process, from government level to humanitarian responses. Static scales can be particularly useful by focusing on broad concepts of development, preparedness and response which should be related to handling many types of crisis effectively. They are also advantageous for relying mostly on static indicators (rarely reestimated or measured by some countries), especially since collecting data at a global scale can be a very troublesome process. However, these scales are mostly conceptually modelled, and since empirical evidence of its correspondence to reality is hard to obtain, they must be used with caution. The COVID-19 pandemic itself has promoted the release of more open data. Nonetheless, data collection procedures can be very inconsistent across the world, meaning that naively using these data to measure impacts may be misleading. Excess mortality can be a more robust metric to these problems, accounting for both direct and indirect deaths caused by the pandemic.

In this chapter, an analysis of the **relation between open-source static risk scales and the COVID-19 pandemic's impact in terms of excess mortality** will be conducted. All countries for which mortality data were available in *FAIRisk* for at least 9 months following the date of the first confirmed case of COVID-19 were considered (41 countries[1]). Since not all of these countries were available in ReadyScore Map, this scale was not considered. This analysis aims t**o provide insights about the possibility of relying on such scales to foresee the damaging impact of a certain crisis – epidemic crisis, in this case – across countries**.

2.1. Excess mortality over different pandemic periods

Studying the evolution of excess mortality after the appearance of COVID-19 in each country can deepen the understanding of its country-level impact in different pandemic periods. Figure 1 exhibits the correlation of countries' excess mortality (P-score) per month. Monthly periods are considered starting from the date of the first confirmed case of COVID-19 for each country. Data from each period was scaled to a 0 to 1 range.

By analyzing the correlation matrix, one can observe that excess mortality rates per country were not consistent at all periods following the start of the pandemic. A strong correlation between all periods would be expected if the country-level impact in terms of excess mortality was consistent despite the time. However, such a pattern is not observed. The first 3 months, for example, exhibit a very weak correlation with the 3 last months under analysis. Another relevant aspect of this inspection concerns the evolution of the pattern of excess mortality across countries over time and its correlation with the overall excess mortality in the 9-month period. As time increases, so does the correlation between monthly and overall excess mortality across countries. This points to a stabilization of the most and least affected countries, following the period of adaptation to the crisis, and can be related to countries' coping capacity at the infrastructure level.

1 Armenia, Australia, Austria, Belgium, Bulgaria, Canada, Chile, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Israel, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Netherlands, New Zealand, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, South Korea, Spain, Sweden, Switzerland, Taiwan, United Kingdom, United States.

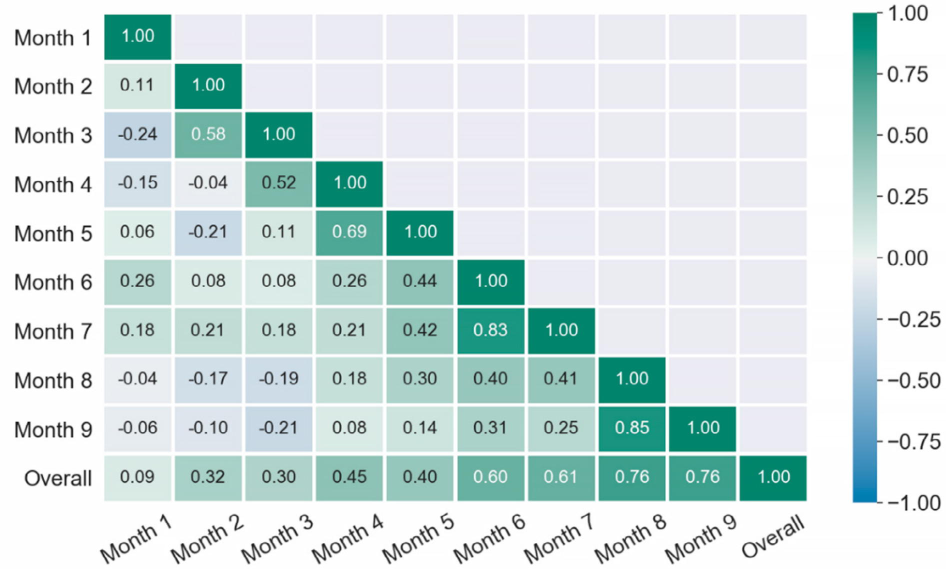


Fig 1: Correlation matrix of excess mortality P-score for each month after the date of first confirmed COVID-19 case for each analysed country (over a 9-month period). Overall refers to the full 9-month period.

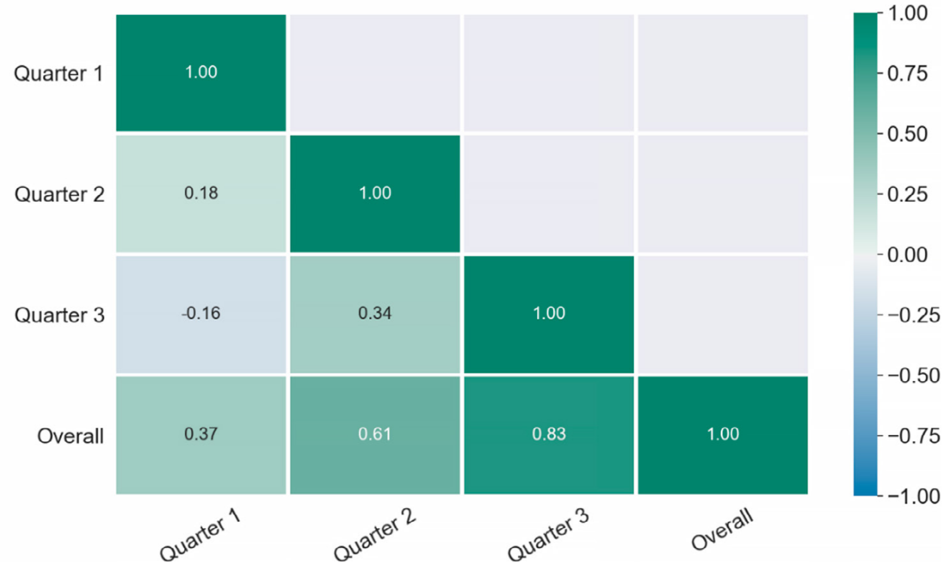


Fig 2: Correlation matrix of excess mortality P-score for each quarter (3-months period) after date of first confirmed COVID-19 case for each analysed country (over 3 quarters). Overall refers to the full 9-month period.

Equivalent analyses can also be conducted from the quarterly breakdown for the same 9-month period shown in Figure 2 Quarter periods will also be considered in the following analysis for concision and readability purposes.

2.2. Correlation with static indexes

2.2.1. INFORM Risk vs. excess mortality

The INFORM Risk model (Marin-Ferrer et al., 2017) relies on the definition of three dimensions – Hazard & Exposure, Vulnerability, Lack of Coping Capacity. Hazard & Exposure relates to the burden that a certain natural or human-induced hazard can be for the exposed community. Vulnerability is the result of the aggregation of two categories: socio-economic vulnerability and vulnerable groups. The analysis is done through hazard-independent indicators but intends to reflect the damaging impact of a hazard in social, political, and economic conditions that can affect the exposed population. The coping capacity dimension addresses the ability of countries to respond to hazards in a formal, organized way, which includes the existence of Disaster Risk Reduction (DRR) programs for mitigation and preparedness/early warning phases at an institutional level and the capacity for emergency response and recovery at an infrastructure level.

Two adaptations of this risk index were proposed. The first focused on the incorporation of epidemics risk in INFORM's Global Risk Index (GRI) while preserving the integrity of the proposed model (Poljansek et al., 2018). Later, in 2020, an experimental adaptation of the INFORM Epidemic Risk Index was created to identify countries most at risk following the impacts of COVID-19 (Poljansek et al., 2020). Figure 3 exhibits the correlation between the final Epidemics and COVID-19 risk indexes and contributing dimensions of INFORM's model and the excess mortality in the 9-month period following the appearance of COVID-19 cases in each country for both of these adaptations.

Given the different patterns of most/least affected countries at each pandemic stage (see section 2.1) and the fact that both models under analysis are static, an uneven distribution of correlation values between pandemic periods and the risk index and dimensions was expected. This distribution can provide relevant information about the strengths and limitations of the models at foreseeing risk at different pandemic stages.

The Hazard & Exposure dimension shows significant differences between the Epidemic and COVID-19 models, especially since the latter only considers person-to-person transmission-related indicators. These indicators appear to show some correlation with excess mortality over the first quarter, but this correlation tends to fade over time.

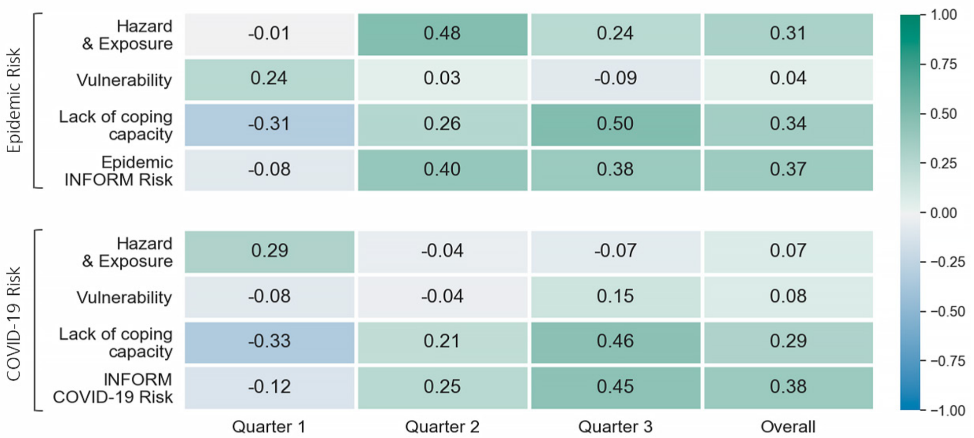


Fig 3: Correlation between excess mortality P-score for a 9-month period after the date of first confirmed COVID-19 case by country (quarters, overall) and INFORM Epidemic and COVID-19 dimensions and risk index.

Vulnerability estimations also vary significantly from the epidemics model to the specific COVID-19 one. COVID-19 Vulnerability solely considers movement, behavioral, and demographics and co-morbidities indicators, which appear to correlate more with this pandemic's impact over time. On the other hand, the broader indicators used to estimate vulnerability in the epidemics scale show a stronger correlation with the first months of the pandemic.

Excess mortality over the first pandemic quarter shows a negative correlation with the Lack of Coping Capacity dimension. If we consider this first stage as the one which should relate the most with countries' preparedness and early warning, this fact may sound surprising as this dimension is indeed the one which considers such indicators. However, the same dimension also considers the capacity for emergency response and recovery, which may be the reason why the correlation with excess mortality increases over time.

All in all, the Epidemic INFORM Risk and INFORM COVID-19 Risk indexes show a similar correlation with the overall excess mortality P-score over the entire 9-month period. The tendency to show a stronger correlation with pandemics' impact on the loss of lives over time is also common. As for the first quarter, one can verify a very week (negative) correlation with both indexes. Moreover, the dimension that correlates the most with this period is different for both scales. As such, while these conceptually modelled scales appear to convey relevant information towards risk modelling, there may be a limitation related with coherently foreseeing risk over the first pandemic months (namely, if we consider the overall balances). Additionally, there may be room for improvement of both models, as risk indexes and overall excess mortality only verify a mild correlation.

2.2.2. GHS Index vs. excess mortality

The GHS Index framework consisted of the application of 140 questions, organized across 85 subindicators, 34 indicators and 6 categories, that aimed to assess countries' capacity to prevent and mitigate epidemics/pandemics. Category names are mostly self-explanatory, but a more thorough description can be found in the 2019 GHS Index report[2] :

A. **Prevention** – level of prevention of the emergency or release of pathogens;
B. **Detection and Reporting** – level of early detection and reporting for epidemics of potential international concern;
C. **Rapid Response** – level of rapid response to and mitigation of the spread of an epidemic;
D. **Health System** – level of health system robustness to trat the sick and protect health workers;
E. **Compliance with global norms** – commitments to improving national capacity, financing plans, and adhering to global norms;
F. **Risk Environment** – level of overall risk environment and country vulnerability to biological threats.

2 https://www.ghsindex.org/report-model/

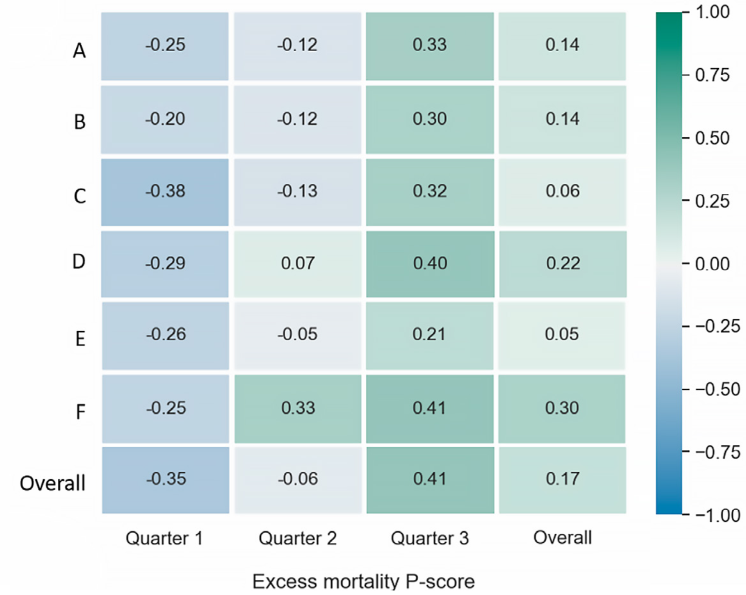


Fig 4: Correlation between excess mortality P-score for a 9-month period after the date of first confirmed COVID-19 case by country (quarters, overall) and (inverted) GHS Index categories and overall score (rows). Legend: [A] Prevention; [B] Detection and reporting; [C] Rapid response; [D] Health System; [E] Compliance with International Norms; [F] Risk environment.

Figure 4 shows the correlation between the score of each category and the overall index and excess mortality over different pandemic periods. GHS scores were inverted to provide a risk-oriented metric instead of a security assessment, using the maximum value of the scale (100) to which each score value was subtracted. Values of the countries under analysis were then scaled to a 0 to 1 range.

While some categories may not have a direct relation with what happens following the first appearance of COVID-19 cases in the community (such as prevention or even detection and reporting), others withhold important information related to handling the spread and response to the disease. Maybe to some surprise, these categories show a mild negative correlation with the excess mortality over the first quarter after the detection of the first case, including the rapid response category. This may be significant by suggesting that particular effort should be employed in understanding how to improve this score. Health system, compliance to norms, and risk environment also may require some improvement. Its correlation with the overall impact in loss of lives is still weak, even though it seems to become stronger over time. This leads to an overall GHS Index (inverted) score that shows a very weak correlation with these 9-months excess mortality, and supports that there may be room for improvement, or even to rethink, the index. This conclusion is also supported by other recent works (Abbey et al., 2020; Aitken et al., 2020).

2.3. Highlights

The relation between the outcome of open-source scales of country risk ranking and COVID-19's impact on excess mortality across countries was explored. The analysis unveiled that modelling the evolution of a crisis and the risk level that it poses at each stage is not a straightforward task, as countries most affected over the first pandemic months were not the same as the ones that accounted for higher overall excess mortality over the 9 months following the break of the disease in each country. This dynamic might be difficult to model with resort to static risk ranks, despite their consideration of preparedness and rapid response indicators. The main limitation of the conducted experiments lies on the fact that only countries for which there was mortality data available (41 countries) could be used for analysis, most of which were OECD members.

- The countries most and least affected by the COVID-19 pandemic varied with its evolution over time;
- Current risk ranking scales still have room for improvement, as these solely verify a mild correlation with the overall impact in loss of lives across countries;
- None of the analysed scales and dimensions/categories were able to foresee the damaging impact of the first quarter for the countries under analysis, despite considering indicators of rapid response in their modelling strategy.

3. Can We Improve Conceptually Modelled Risk Scales With Data-Driven Insights?

As we have seen, creating a globally available risk scale is a demanding task. Even considering the modest number of 41 countries, it was clear that there might be some room for improvement of such scales, presumably not only to foresee the impact of a specific disease, such as COVID-19, but for broader hazards. The analysed risk models were created by specialists of a variety of fields related with disaster management and translate conceptual approaches for risk modelling. But are there hidden patterns in the data that can be useful for improving these models?

In this chapter, methods of unsupervised learning will be used to **extrapolate possible underlying relations between hazard-agnostic countries' indicators**. Then, excess mortality over time per country will be studied within each cluster to **verify whether there are similarities between how countries of the same cluster were affected by the pandemic**.

3.1. Unsupervised modelling

Data from two categories of *FAIRisk* repository's data model – indicators and scores - were used to conduct a global analysis that aimed to find hidden relations between countries by modelling the discriminant power of their attributes. We considered all attributes available for more than 100 countries. Countries that only had data for less than 80% of the considered attributes were excluded. This resulted in a total of 184 countries and 177 attributes for analysis. All data were normalized using *FAIRisk* repository's methods for scores and indicators normalization (min-max scaling).

The 41 countries for which there were mortality data available using *FAIRisk* resources for at least 9 months after the appearance of the first COVID-19 case were selected from this dataset. This selection was required to enable a posterior analysis of the hypothesized relations or similarities of excess mortality evolution within clusters. Nonetheless, and to enable other analysis of potential relevance, the modelling steps described hereon were also mimicked for the entire set of 184 countries. All figures derived from this global analysis are analogous to Figures 5 to 8 and are available in Annex (A.1).

3.1.1. Principal Component Analysis

Principal component analysis (PCA) was employed to expose patterns in the multivariate dataset. PCA facilitates a description of the relationship between variables while attempting to explain the total variation of the data. Figure 5 represents the cumulative variance explained by each consecutive principal component. One can observe that the first component alone explains over 20% of the variability of the entire set. This component combines 8 variables, namely anti-tobacco mass media campaigns, current health expenditure per capita, public trust in politicians, International Health Regulations (IHR) core capacity of human resources, and the average price of a 500 ml beer.

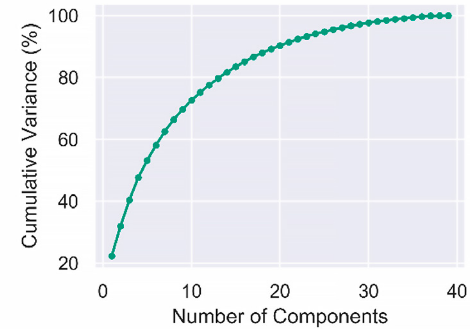
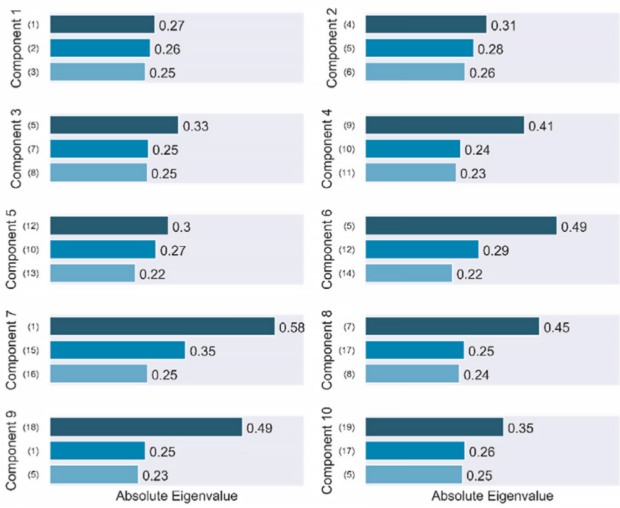


Fig 5: Cumulative variance according to the number of principal components.

Figure 6 discloses the 19 attributes that contribute the most towards the 10 components, which represent over 70% of the overall variance, for transparency purposes. These attributes concern several societal dimensions which may or may not be explicitly associated with communities' risk. However, their combination can be indicative of a profile that can disclose hidden relations between countries by simultaneously accounting for societal, behavioural, lifestyle, and other indicators that contribute to countries' identity.



1. Anti-tobacco mass media campaigns
2. Current health expenditure (CHE) per capita in US$
3. Public trust in politicians
4. Enforcing bans on tobacco advertising, promotion and sponsorship
5. National smoking ban: number of places smoke-free
6. Warning about the dangers of tobacco
7. IHR core capacity of human resources
8. IHR capacity score: Points of entry
9. International tourism, number of arrivals
10. Treatment success rate: new TB cases
11. Number of vets
12. Treatment success rate: HIV-positive TB cases
13. Average price 500 mls Beer in US$
14. Surveillance
15. Previously treated cases tested for RR-/MDR-TB (%)
16. TB patients with known HIV status (%)
17. Nursing and midwifery personnel (per 10,000)
18. Treatment success rate: previously treated TB cases
19. Psychiatrists working in mental health sector (per 100,000)

Fig 6: Barplot of the 10 components (over 70% explained variance) and the respective eigenvalues of the 3 indicators with more weight. A short description of each indicator is also presented (right).

3.1.2. *K*-means clustering

A total of 26 components, explaining over 95% of the variance, were then used to feed a *K*-Means clustering algorithm. *K*-Means assigns samples to a cluster based on the closest cluster center according to a Euclidean distance function. The *K* variable should be predefined (number of clusters), even though the calculation of its optimal value is not straightforward. Some methods, such as the elbow or the average silhouette, may assist this process. Using the elbow method, the optimal *K* should consist of the point of greatest decrease in within cluster sum of squares (WCSS) point. The average silhouette method considers that the maximum average silhouette should provide the optimal number of clusters. From the analysis of Figure 7, one can verify that both methods seem to point to an optimal *K*=2.

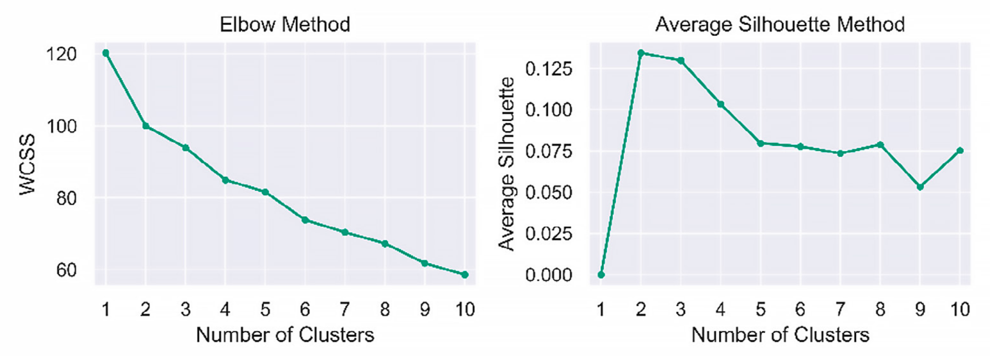


Fig 7: Results of elbow (left) and average silhouette (right) methods.

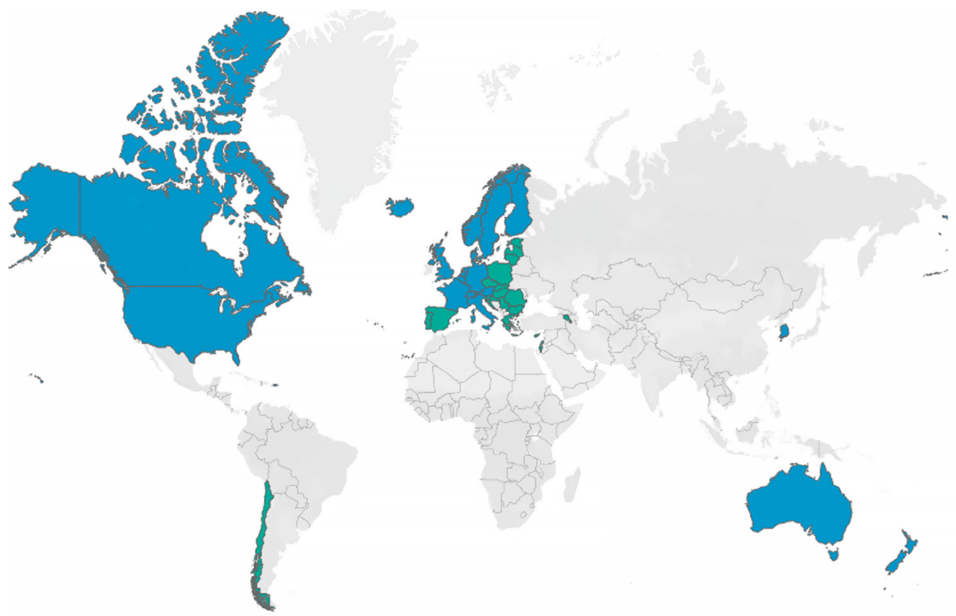


Fig 8: Worldmap representation of the 2 clusters of countries under analysis, Cluster A (blue) and Cluster B (green).

Figure 8 exhibits the outcome of the clustering algorithm (considering 2 clusters):

> ***Cluster A***: *Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Italy, Luxembourg, Netherlands, New Zealand, Norway, South Korea, Sweden, Switzerland, United Kingdom, United States*
> ***Cluster B***: *Armenia, Bulgaria, Chile, Croatia, Cyprus, Czech Republic, Estonia, Greece, Hungary, Israel, Latvia, Lithuania, Malta, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain*

The worldmap representation of the clusters indicates that Cluster A accounts for a considerably larger area, presumably also with lower population density, while Cluster B seems to be more self-contained. The variability of communities belonging to Cluster A is thus expected to be higher than that of Cluster B. The relevance of this outcome will also be studied in the next subsection.

3.1.3. Relationship and similarity of excess mortality within clusters

A set of experiments was designed to study the relationship and similarities between excess mortality over time for countries belonging to the same cluster. Two different analyses were conducted by computing Pearson correlations and Dynamic Time Warping (DTW) distances between time series.

Correlation

Figure 9 shows the correlation matrixes of each cluster, using countries' monthly series of excess mortality P-score. An analogous figure showing the correlation results for all countries is also available in Annex (A.2). Despite the hazard-independent country clustering methodology, one can promptly observe the strong correlation between excess mortality evolution of most countries of Cluster B (with the exception of Chile, Cyprus, Israel, Serbia, and Spain). The same conclusion is not as clear for countries of Cluster A.

The boxplot of Figure 10 can help us to deepen this analysis by disclosing the distribution of correlation values between the excess mortality evolution of each country and others of the same cluster vs. all available countries. Generally, the results show that the distribution of correlation results for all available countries is sparser than within clusters. Once again, this conclusion is particularly supported by the results related to Cluster B, which was seemingly expected due to the lower community variability of this cluster when compared to Cluster A.

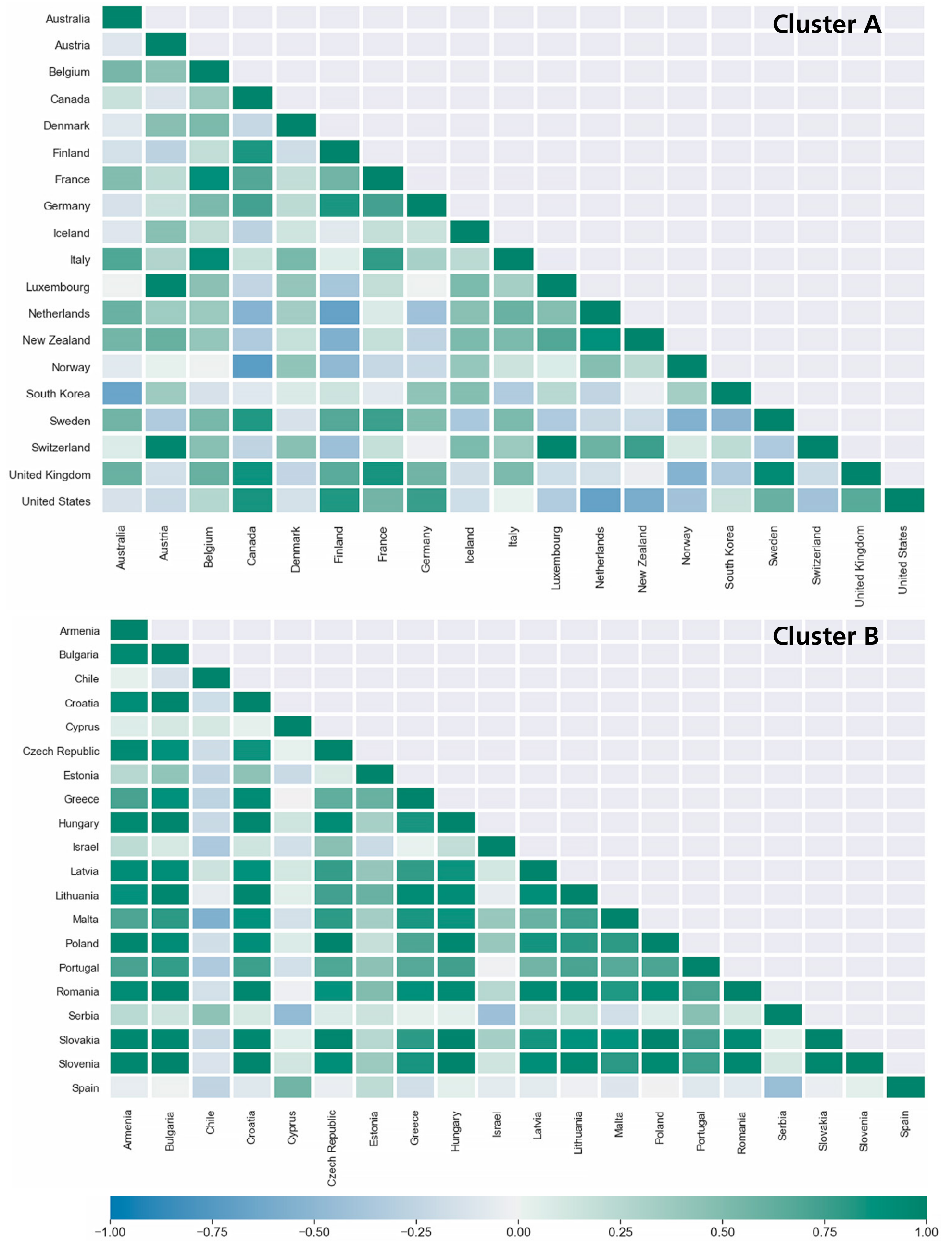


Fig 9: Correlation between countries of clusters A (top) and B (bottom), using monthly series of P-score excess mortality per country as input.

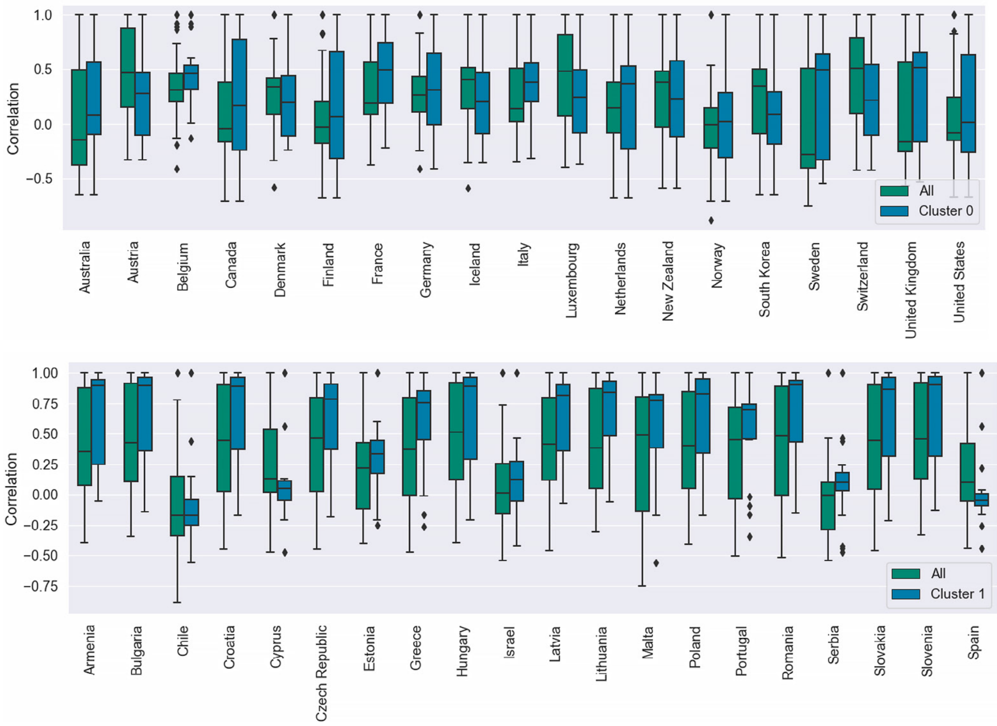


Fig 10: Boxplot chart for the results of correlation between the monthly series of excess mortality of a country vs. other countries of its cluster or all available countries.

DTW

DTW is a time series similarity measure. In contrast to other approaches, like the Euclidean distance, this distance disregards signals' phase, which makes it adequate to compare temporal sequences without prior knowledge of whether they vary in speed. Interpreting this metric is mostly straightforward - lower distances between time series support higher similarity. Figure 11 shows the boxplot chart that summarizes the results of the comparison between each countries' monthly evolution of excess mortality and that of other countries of the same cluster vs. all available countries.

By analysing the results, one can generally verify a tendency for smaller distances between time series of countries belonging to the same cluster in comparison to all countries' outcome. This conclusion is mostly consistent for both clusters, despite some exceptions (e.g. Germany, Austria). This constitutes another evidence that the hazard-agnostic clustering approaches can disclose relevant hidden patterns in the data which can assist the discovery of affinities between countries and, ultimately, have the potential to improve current risk modelling strategies with purely data-driven insights.

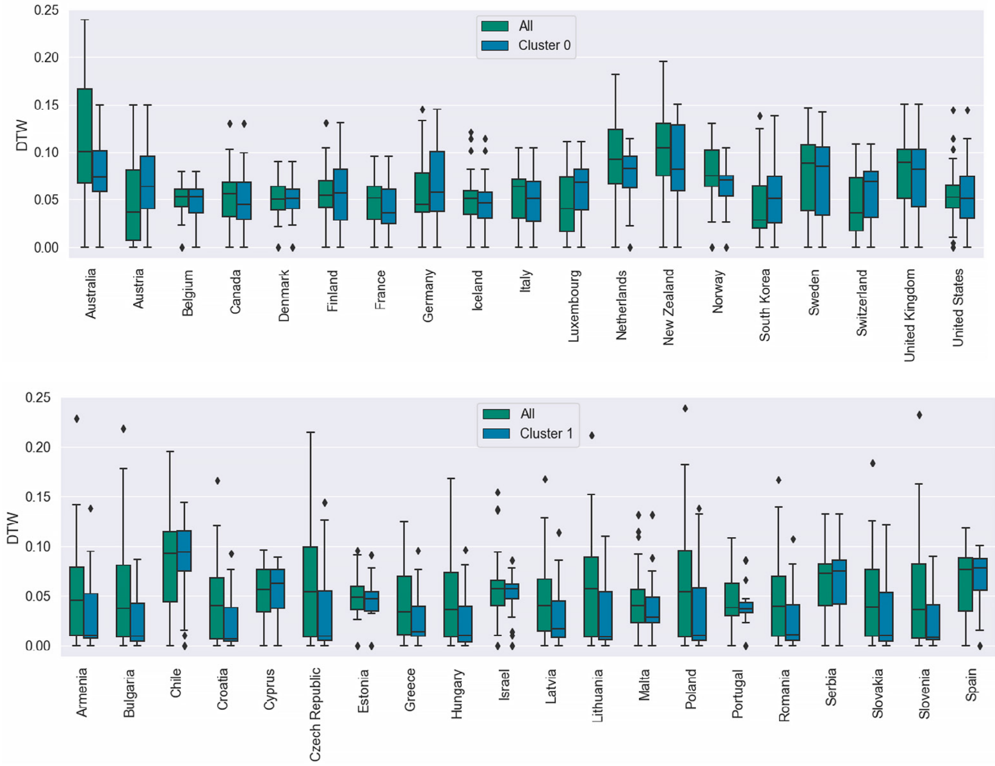


Fig 11: Boxplot for the results of DTW between the monthly series of excess mortality of a country vs. other countries of its cluster or all available countries.

3.2. Highlights

The implemented unsupervised learning methods revealed underlying relations between behavioral, lifestyle, societal and health-related indicators available for the countries under analysis. These countries were then grouped in clusters using the principal components that explained over 95% of all data variance. Despite the fully hazard-agnostic modelling approach, our results support the similarities between the evolution of excess mortality over time of countries belonging to the same cluster. This outcome indicates that the hidden relation between countries' attributes might withhold important information that can assist risk modelling strategies, despite its risk-blind approach.

- Behavioral, lifestyle, societal, and other country-level indicators can be blindly modelled to expose groups of countries which appear to show similar excess mortality evolution over time following the appearance of COVID-19;
- Hidden relations between countries' attributes may withhold important and purely data-driven information with potential to assist conceptual risk modelling expertise.


AMPULAB
COAGULATION
Цитрат натрия
3.8%
Кат. № 302112
4
ml
REF13030
K3EDTA 7.2 mg
SOYAGREENTEC

4. Can Open Data Be Used To Forecast Excess Mortality At A Global Scale?

During a pandemic crisis, the need for accurate and continuous data is of utmost importance to evaluate current containment measures and provide assistance in the creation of new responses upon forecasted outcomes. Therefore, decision-makers can benefit from larger datasets to study the variety of aspects that concern the situation at hands. Since the excess mortality poses as an important indicator of countries' response to the COVID-19 pandemic, its precise forecasting shall be a powerful aid to the decision-making process, enabling the understanding of how some measures may impact future outcomes.

In light of this, we evaluated how **open data resources can assist a timely excess mortality estimation for different countries**. As such, we explored the information gathered using *FAIRisk* resources to understand how continuous variables, namely the COVID-19-specific and population mobility indicators, relate with the confirmed excess mortality. Afterwards, using both the aforementioned continuous parameters and the scores and indicators that characterize each country, we applied Machine Learning techniques to forecast the excess mortality for different countries.

4.1. Modelling post-COVID-19 open data using excess mortality as target

Regression techniques were used to gather more insights into the relationship between post-COVID-19 available open data and excess mortality. All countries for which there were mortality data available in *FAIRisk's* mortality category were considered for this analysis. The group of attributes selected was provided by Our World in Data COVID-19 and the Movement Range Maps datasets using *FAIRisk* resources. The two datasets provide data on confirmed cases, deaths, hospitalizations, testing, and metrics reporting how populations respond to physical distancing measures.

A Linear Regression model was selected to facilitate the interpretation of the learned relationship's linearity, which is not always as evident for other approaches. We applied Linear Regression using Ordinary Least Squares to each available country. Attributes' importance was evaluated by the absolute value of its *t*-statistic:

$$\left|t_{\widehat{\beta_J}}\right| = \frac{\widehat{\beta_J}}{SE(\widehat{\beta_J})}$$

where $\widehat{\beta_J}$ represent the learned attribute coefficients and SE the regression standard error. The Log-transform was also used to approximate the target data to the normal distribution.

The R^2 for all countries was 0.95 ± 0.09. The most important attributes are summarized in Figure 12. The indicators related to mobility were deemed the most relevant, in particular, the mean StayPut and Change, which measure the proportion of users staying put within a single location and change in movement relative to baseline, respectively. These attributes are strongly correlated with the stringency measures imposed during the pandemics.

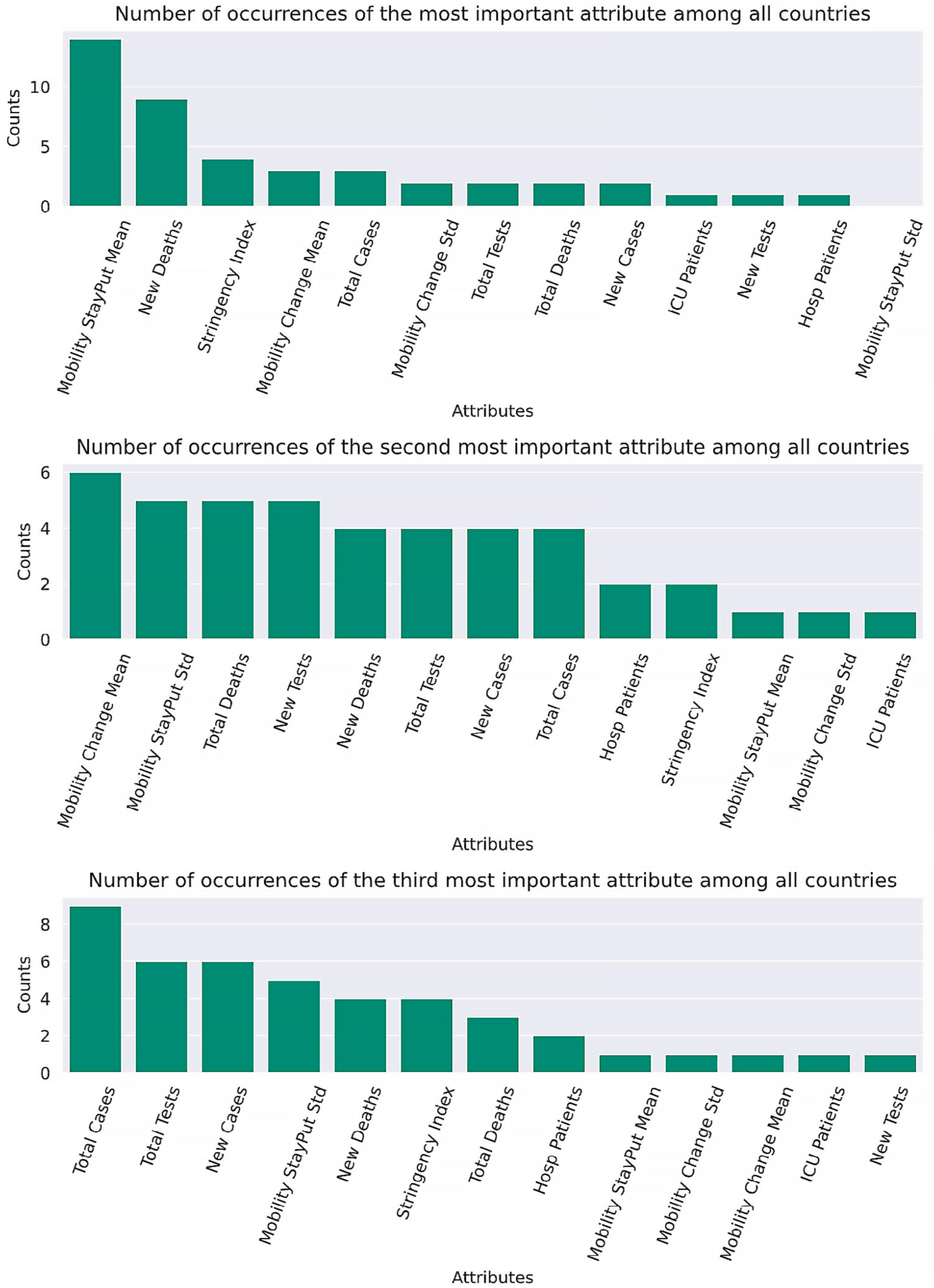


Fig 12: Counting of the number of occurrences of attributes among all countries, where a given attribute was weighted as the most important (top), second most important (middle), and third most import (bottom).

Despite these data-driven insights might contribute to improve conceptually modelled static risk scales, some considerations must be taken into account in their interpretation. Linear regression assumes that each instance is independent, which might not always hold true in this scenario. While linear effects might be easy to quantify, they are not adequate to measure attribute interactions or non-linearities. Future work might consist of using additional methods, such as regression splines or mixed effect models.

4.2. Excess mortality forecasting

The precise forecasting of excess mortality is a key task to understand how the current pandemic status in a given country will affect the population mortality in the future, allowing decision-makers to launch new containment measures, or, on the other hand, to lift some restrictions.

Forecasting techniques have been applied in different contexts, such as inventory management or stock price prediction, often leveraging past and present values of the target variable to predict its value in the future (Hyndman & Athanasopoulos, 2018). However, foreseeing the future is not an easy task, especially in complex topics, where the mechanism that affects some target is not known, or it depends on a large set of variables, if not both. The factors that directly impact the excess mortality in some country when facing a pandemic are no exception.

Hence, estimating the excess mortality in such extraordinary times is a hard task, and most attempts end up failing, especially when the forecasting window is large – often more than a week (Friedman et al., 2020). Such models use simple statistical forecasting techniques and limited sets of variables, which overfit the training set and perform poorly on unseen data. As a result, a universal excess mortality forecasting model is yet to be developed.

Despite the difficulties of this task, there is a consensus that a broad range of factors affects the deaths in excess during pandemic times. The mortality will be impacted not only by the number of new cases, but also by the countries' health infrastructure and economic status, to name a few. Therefore, the vast data sources gathered with *FAIRisk* resources potentially include most variables that affect mortality in a given country. Combining such data with more robust Machine Learning techniques is expected to improve the accuracy of excess mortality forecasting.

As such, we leveraged these data to train an excess mortality forecasting model, which relies on both daily collected data and fixed indicators that characterize different countries. We used time series data from the COVID-19 and mobility categories, which were interpolated into weekly values, to eliminate daily fluctuations on registered values. Additionally, indicators and scores were used so that models could consider the differences between countries, and how some parameters impact mortality. From the demographic information, we retrieved the total population and the age structure of each country, with the share of children (0 to 14 years), active age (15 to 64 years) and elderly (65 plus years). The total population was also used to normalize some absolute variables to a general scale, such as the number of new COVID-19 cases or hospitalized patients. Finally, to provide models a way to consider the phase of the pandemic, a temporal feature was created, which is weekly incremented from the week of the first COVID-19 case in each country. Regarding the target variable, the weekly total excess mortality for each country was computed and the P-score is used by the models.

Data Preparation

Since models require a set of input features, which will be used to predict the future excess mortality, a data preparation process was applied.

Firstly, from the complete set of scores and indicators, with 484 variables at the time of writing, we removed those which are not available for at least 90% of all countries. Then, we dropped all countries which do not have demographic information both from the total population and the three age structures. Selected countries should also have at least 50% of the remaining scores and indicators. This process restricted the static dataset to 44 countries with 194 variables. Then, the continuous data was also processed, and countries with more than 50% of all COVID-19 and mobility variables missing in all days were discarded. At last, only countries with computed excess mortality could be used. These processes resulted in a dataset with 33 countries[1].

The data preparation process highly limited the number of countries that can be used to train Machine Learning models for excess mortality forecasting to the European level. Although future studies can be designed using less restrictive requirements, i.e. assuming that more data may be missing, it probably cannot be done without impairing models' performance. Therefore, these limitations support the need for the release of more open data at a global level.

Definition of Forecasting Windows

To forecast excess mortality, Machine Learning models need a set of past information from a specific time frame, which is expected to trigger the future number of deaths. Since COVID-19 and mobility categories contain weekly sampled parameters from each country, further processing is still necessary to create the samples to be modelled. For this process, we leveraged the rolling mechanism from the TSFRESH python library (Christ et al., 2018), which defines forecasting windows from time series. This tool requires the setting of the expected window size to be obtained, which we defined through three parameters:

- Observation window: time frame from which continuous values were extracted to be input to the models.
- Interval window: a time frame to be ignored, in case we aim at predicting future values that do not immediately follow the observation time frame.
- Target window: the time frame from which the excess mortality will be forecasted.

Figure 13 illustrates the rolling mechanism, from which we retrieve the samples with respect to the defined observation, interval and target window sizes.

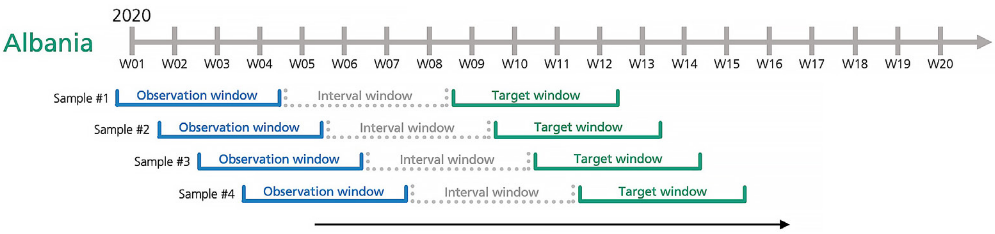


Fig 13: Scheme of samples' retrieval with the rolling mechanism.

Depending on the size of the forecasting window, data from several weeks will be retrieved. Instead of using all data, we extract features from such data. Therefore, we compute the minimum, maximum, mean, median and standard deviation, while we also keep the most recent value from each parameter. Finally, the static information corresponding to each sample's country was concatenated into the defined feature set.

1 Albania, Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Iceland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom.

Definition of Train and Test sets

Before forecasting the excess mortality, the dataset must be split into two groups, one for training Machine Learning models and the other for testing the performance against new data, to avoid overfitting. Therefore, we used 80% of all samples for training and the remaining 20% for testing. Considering that models may overfit to the countries used in training and that the different waves of the pandemic impacted the excess mortality differently, we did not sample training samples randomly from the complete dataset, neither did we use the initial 80% of data for the training. Instead, we selected 80% of the countries for training, leaving the remaining 20% for testing.

Before modelling, we performed a final processing stage where we normalized all features using their minimum and maximum (defined from the train set and applied to both). Quasi-constant features (in more than 99% of train set values) and those which were highly correlated (over 90% between train set values) were removed. Finally, we imputed all missing data with the -1 value, to contrast with the remaining values (between 0 and 1).

Different observation window sizes will produce distinct features vector shapes. For instance, if the observation window includes four weeks of data, then the statistics computed for each continuous attribute (from COVID and mobility categories) will have meaningful values. On the other hand, if the observation window is of one week, then the statistics for each attribute will be retrieved from the observation week's single value, creating either repeated (mean, median, minimum, maximum and the most recent) or constant (standard deviation) features. Therefore, in this case, all statistics but one are dropped from the final feature set.

Regarding the forecasting task, two different approaches were considered:

- Regression: forecast absolute excess mortality values in %.
- Classification: forecast excess mortality as groups.

4.2.1. Continuous target

The forecasting of absolute values of excess mortality is defined as a regression task, where the target is not a category nor a binary, but a continuous value. For this purpose, we leveraged the TPOT python API, an automated Machine Learning tool which uses genetic programming to test different pipelines, from the features level to the model selection and parameter optimization processes to select the best Machine Learning approach for each particular objective (Olson et al., 2016).

Considering the need to avoid overfitting to the countries in the training set, we leveraged the Leave-One-Out Cross-Validation strategy, where the models are trained in several runs, each training the model with all train set countries but one, which is used for validation (Webb et al., 2011). After this process, the best pipeline is selected and used as the final model. Then, to understand how the model behaves in unseen data, we tested its performance against the test set using three metrics:

- Mean absolute error (MAE);
- Mean squared error (MSE);
- Coefficient of determination (R2).

Training Procedure

Besides optimizing the Machine Learning pipeline to obtain an accurate regression model both in the train and test sets, it is important to find the best observation, interval and target windows. Thus, several different combinations of window sizes were tested:

- Observation window size – 1, 2, 4 and 8 weeks;
- Interval window size – 0, 1, 2 and 4 weeks;
- Target window size – 1, 2 and 4 weeks.

We also asserted the models' performance for these combinations without the created temporal feature, and by applying the Log-transform to approximate the target with the normal distribution.

Results

After running the TPOT optimization pipeline with the different combinations of observation, interval and target windows and using the aforementioned variations, the best pipeline regarding the MAE on the test set was achieved using an observation window of 1 week and a target window of 4 weeks, without any interval week. In practice, this estimation is done using the values of a single week to forecast the mean excess mortality of the following month. The achieved results are described in Table 2.

Table 2 Performance metrics of the best regression model (Random Forest Regressor).

R^2	Mean Absolute Error	Mean Squared Error
0.36	**8.58**	**217.93**

The best pipeline used a Random Forest Regressor with the temporal feature and the Log-transformed target. The results point towards the feasibility of the method, with a MAE of 8.58% for the forecasted excess mortality on the test set. However, the R^2 indicates that the model can only explain 0.36 of the variance of the target variable, which may imply that the model is not reliable enough to be used in a real-world, unsupervised application. Evidently, this also means that there is room for improvement of the current methodology.

The 10 most important features for the forecasting methodology are depicted in Figure 14. COVID-19-related features reveal a considerable impact on models' performance, with the test positivity rate and the number of new cases standing out of the remaining set of features. The temporal feature that provides the model with a sense of the pandemic phase of the current prediction also has an important role in forecasting.

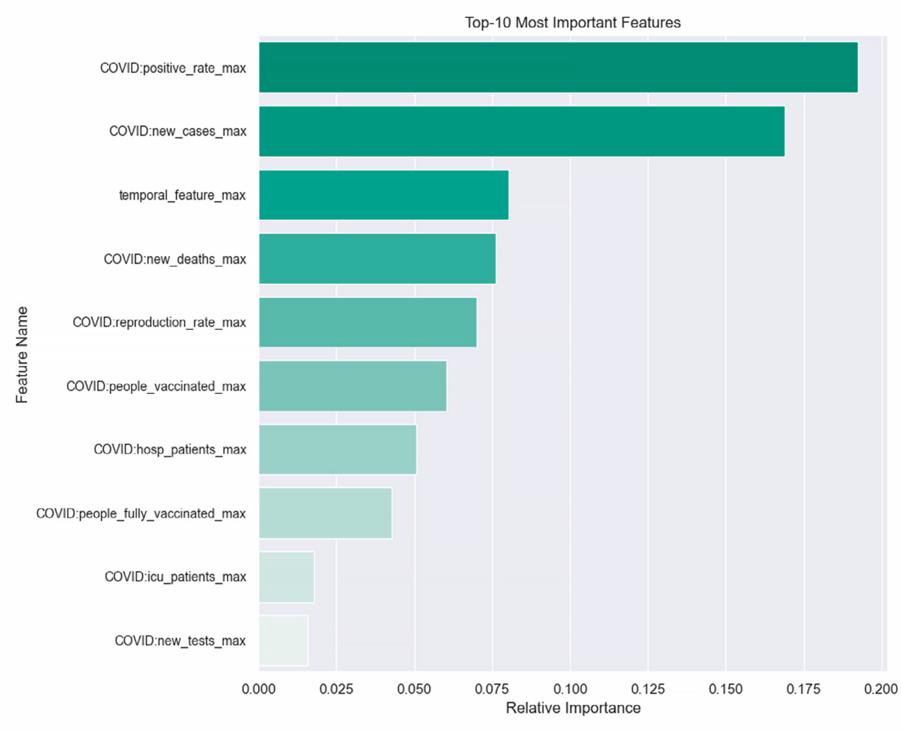


Fig 14: Ten most important features for the selected regression model for excess mortality forecasting.

4.2.2. Discrete target

Despite the added value of forecasting excess mortality as a continuous target, as it allows estimating absolute mortality values, this endeavour may simply be too ambitious using the currently available data with approaches of modest resource consumption. Bearing this in mind, we proceeded to a different approach, where intervals of excess mortality were used as the target instead of continuous values.

Inspired in the groups defined by the "Coronavirus Excess Deaths Tracker" from the Economist and considering the excess mortality distribution from the 33 used countries, the following groups of excess mortality categories were used: <0%; 0-25%; 25-50% and >50%. Figure 15 represents the distribution of all samples over each category of the train set. This representation indicates that the dataset is unbalanced towards the categories that consider up to 25% of excess mortality. The most critical periods concerning excess mortality are scarcer, and are typically associated with peaks of pandemic waves for the different countries.

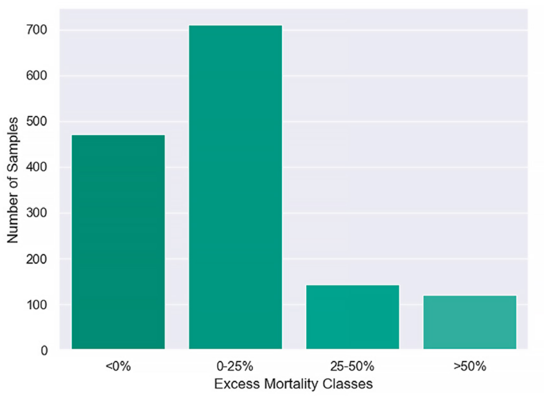


Fig 15: Distribution of the train set samples, according to the defined categories (i.e., classes).

In this task, we took a similar extensive modelling approach as before, using the TPOT optimization pipeline (Olson et al., 2016). Different observation, interval and target window sizes were tested to find their optimal values. The impact of the temporal feature was also studied.

The evaluation metrics, which are necessarily different in a classification task, used to evaluate pipelines' performance for the test set were the following:

- Weighted F1-score;
- Accuracy;
- Weighted precision;
- Weighted recall.

Results

After running the TPOT classification pipeline using the previous set of window sizes, we selected the best excess mortality classifier based on the maximum weighted F1-score attained on the test set. The pipeline which relied on an observation window of 1 week and a target window of 1 week, without any interval, attained the best results. The selected model was a Random Forest Classifier, which considered the temporal feature. Its performance on the test set is available in Table 3.

Table 3 Performance metrics of the best regression model (Random Forest Regressor).

Weighted F1-Score	Accuracy	Weighted Precision	Weighted Recall
65.59%	66.50%	68.46%	66.50%

The Random Forest attained the highest F1-score in the test set (~66%) across the universe of experiments with all pipelines and all different window sizes. Figure 16. depicts the normalized confusion matrix for the test set, with the four considered categories. The matrix shows that most of the confusion occurred within ordinally consecutive classes. This is an important analysis, since this can be considered an ordinal problem, as, for example, the impact of misclassifying >50% samples as <0% can clearly have more troublesome consequences than misclassifying them as 25-50%. As such, while these results could be improved - especially with the release of more open data that can provide more examples of each class (namely, for the least represented ones), the classification of excess mortality with a single model for all countries seems promising, despite our demanding evaluation process which preserved country-independence. Using dedicated ordinal regression techniques can also be an adequate approach to explore as future work.

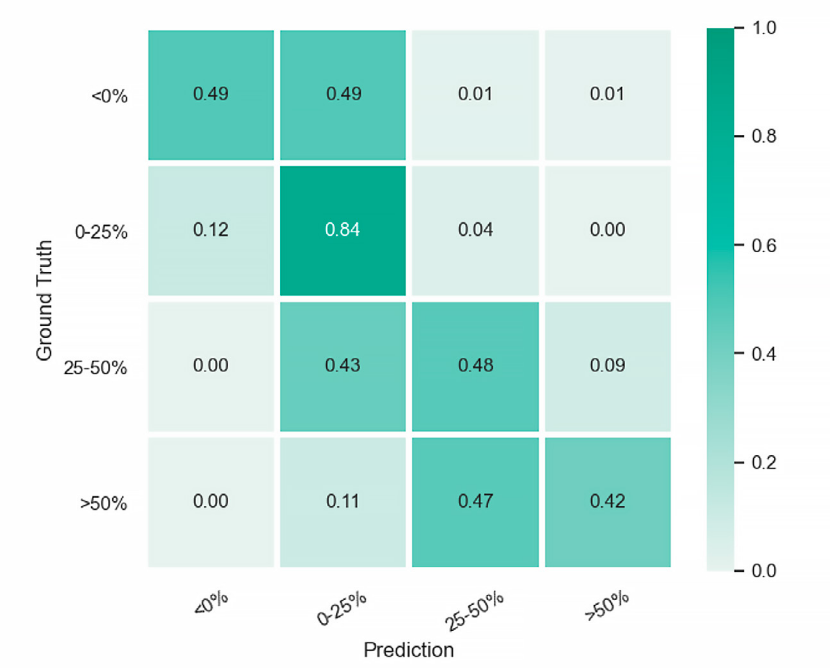


Fig 16: Normalized confusion matrix of the classification results for the test set.

4.3. Highlights

Using mostly interpretable methods to model excess mortality proved to be an overall difficult task. Despite our efforts to understand this target variable, and model it as both continuous and discrete, both forecasting methods are yet to provide highly accurate results. Nevertheless, there may be hope for a single model to provide a timely approximation of excess mortality for several countries at once, by passing a description of such countries to the model with static indicators. This is a promising outcome, and should be explored using more sophisticated forecasting approaches, perhaps combining some of the insights derived from the previous chapter, i.e. taking advantage of the potential of country clustering to improve excess mortality estimations. The main limitation of the work presented relies on the limited number of countries that could be used for analysis. The release of more open data could, therefore, be of utmost importance.

- Using a single model to forecast country-level excess mortality seems to be feasible, but limited by the modest amount of data available;
- COVID-19-related variables, such as the test positivity rate and number of new cases, were the most significant for the forecasting models;
- Forecasting models make better predictions using lower observation timeframes with close targets, i.e. imediatelly consecutive weeks.


COVID-19
Vaccine
2019-nCoV

5. Conclusion

As risk modelling increasingly becomes a widespread priority, so does the role of the data that can contribute towards its success. The work presented in this document can contribute to promote this discussion while encouraging the adoption of adequate open-data and science practices that can boost scientific knowledge and increase the reliability of prospective studies by promoting transparency, interoperability and reusability of resources. Several different approaches of data modelling were explored to showcase the potential of combining more country-level multimodal data to improve current risk modelling strategies, both in hazard-agnostic and hazard-oriented contexts.

Our experiments can provide insights towards deeper discussions about how current risk ranking scales can be improved. They also point to a conclusion that country-level attributes' data withhold hidden relations with the potential to assist conceptual risk modelling expertise and that using a single model to forecast country-level excess mortality is likely to be feasible, despite being extremely limited by the amount of available data.

The preliminary data analysis disclosed in this document can be enriched with further, deeper mining experiments. Future work may include, for example, modelling countries' trends over time instead solely the most recent indicators' data. Another interesting analysis concerns the study of excess mortality by age groups (also available through *FAIRisk* resources), which can unveil important information, especially to study the COVID-19 phenomena across countries, and hopefully contribute towards a better understanding of worldwide patterns. Finally, as new open data repositories are always showing up, new data can be used to extend this analysis (e.g. excess mortality data for more countries).

Whether they are of epidemic or any other nature, worldwide hazards will always be a possibility. This shall compel communities towards their efficient organization and the creation of effective prevention and response plans. *Survival of the fittest* is no longer an individual responsibility. It is a community challenge. And data may as well become the brains of that community.

References

Abbey, E. J., Khalifa, B. A., Oduwole, M. O., Ayeh, S. K., Nudotor, R. D., Salia, E. L., Lasisi, O., Bennett, S., Yusuf, H. E., & Agwu, A. L. (2020). The Global Health Security Index is not predictive of coronavirus pandemic responses among Organization for Economic Cooperation and Development countries. *PloS One*, 15(10), e0239398.

Aitken, T., Chin, K. L., Liew, D., & Ofori-Asenso, R. (2020). Rethinking pandemic preparation: Global Health Security Index (GHSI) is predictive of COVID-19 burden, but in the opposite direction. *The Journal of Infection, 81*(2), 318–356. PubMed. https://doi.org/10.1016/j.jinf.2020.05.001

Aron, J., & Muellbauer, J. (2020). Measuring excess mortality: The case of England during the Covid-19 Pandemic. *INET Oxford Working Paper, 2020–11*. https://www.inet.ox.ac.uk/files/revised-15.54-18-May-20-Aron-Muellbauer-Revised-INET-Excess-Mortality-article-x.pdf

Beaney, T., Clarke, J., Jain, V., Golestaneh, A., Lyons, G., Salman, D., & Majeed, A. (2020). Excess mortality: The gold standard in measuring theimpact of COVID-19 worldwide? *Journal of the Royal Society of Medicine, 113*(9), 329–334. https://doi.org/10.1177/014107682095680

Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). *Neurocomputing*, 307, 72–77.

Friedman, J., Liu, P., Gakidou, E., COVID, I., & Team, M. C. (2020). Predictive performance of international COVID-19 mortality forecasting models. *MedRxiv*.

Ghislandi, S., Muttarak, R., Sauerberg, M., & Scotti, B. (2020). News from the front: Estimation of excess mortality and life expectancy in the major epicenters of the COVID-19 pandemic in Italy. *MedRxiv*.

https://doi.org/10.1101/2020.04.29.20084335

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.

Krelle, H., Barclay, C., & Tallack, C. (2020). *What is the fairest way to compare COVID-19 deaths internationally?* https://www.health.org.uk/news-and-comment/charts-and-infographics/understanding-excess-mortality-the-fairest-way-to-make-international-comparisons

Marin-Ferrer, M., Vernaccini, L., & Poljansek, K. (2017). INFORM Index for Risk Management. *Concept and Methodology Report-Version*.

Olson, R. S., Bartley, N., Urbanowicz, R. J., & Moore, J. H. (2016). Evaluation of a tree-based pipeline optimization tool for automating data science. *Proceedings of the Genetic and Evolutionary Computation Conference 2016*, 485–492.

Poljansek, K., Marin-Ferrer, M., Vernaccini, L., & Messina, L. (2018). Incorporating epidemics risk in the INFORM global risk index. Publications Office of the European Union. https://data.europa.eu/doi/10.2760/647382

Poljansek, K., Vernaccini, L., & Marin-Ferrer, M. (2020). *INFORM Covid-19 Risk Index* [Text]. Publications Office of the European Union. https://ec.europa.eu/jrc/en/publication/annual-reports/inform-covid-19-risk-index

Ritchie, H., Ortiz-Ospina, E., Hasell, J., & Roser, M. (2020). *Coronavirus Pandemic (COVID-19)*. https://ourworldindata.org/coronavirus

Webb, G. I., Sammut, C., Perlich, C., Horváth, T., Wrobel, S., Korb, K. B., Noble, W. S., Leslie, C., Lagoudakis, M. G., & Quadrianto, N. (2011). Leave-one-out cross-validation. *Encyclopedia of Machine Learning*, 600–601.

A. Annex

A.1. Unsupervised modelling: Global analysis

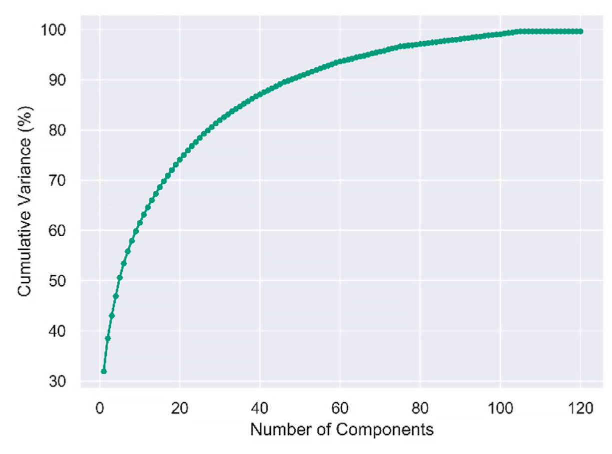


Fig A-1: Cumulative variance according to the number of principal components of the global analysis.

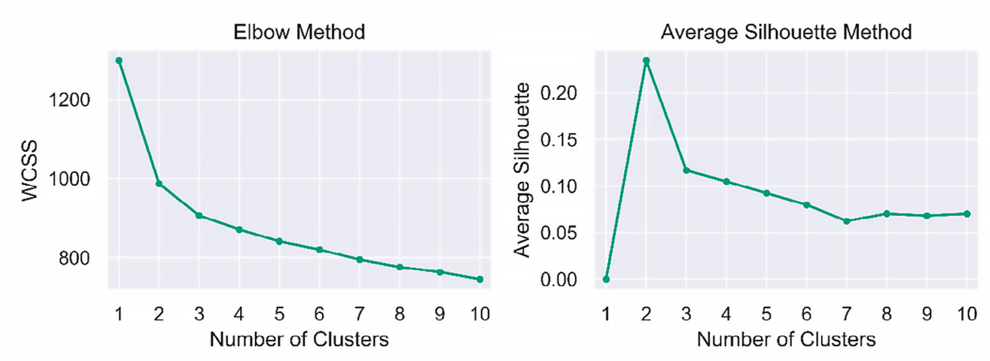


Fig A-2: Results of elbow (left) and average silhouette (right) methods for the global analysis.

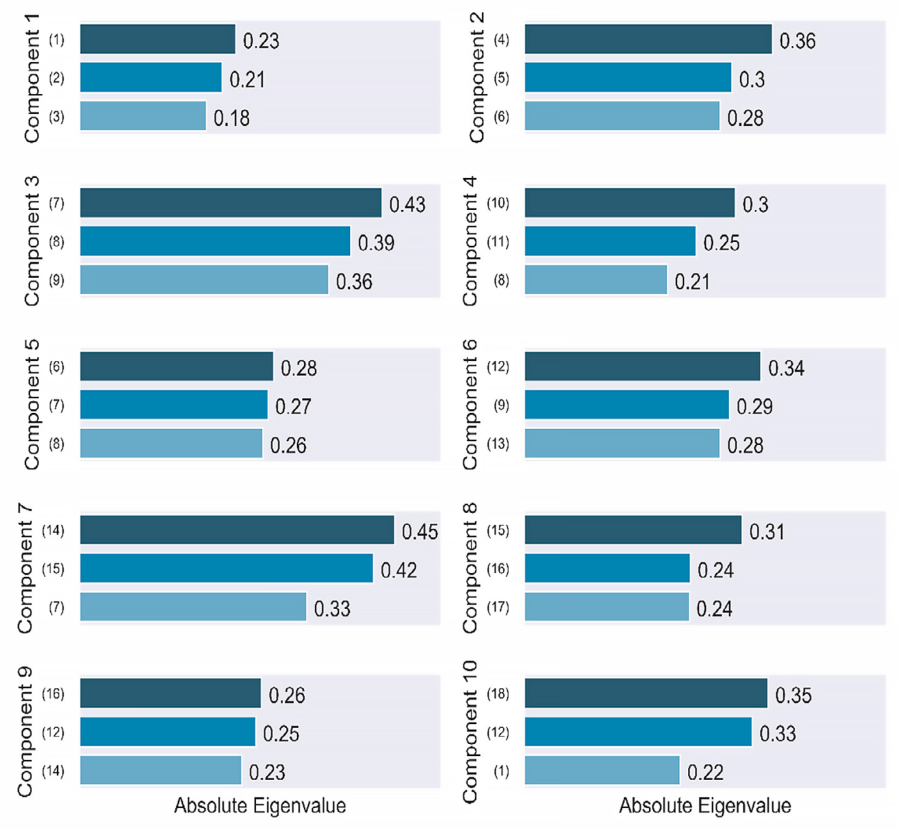
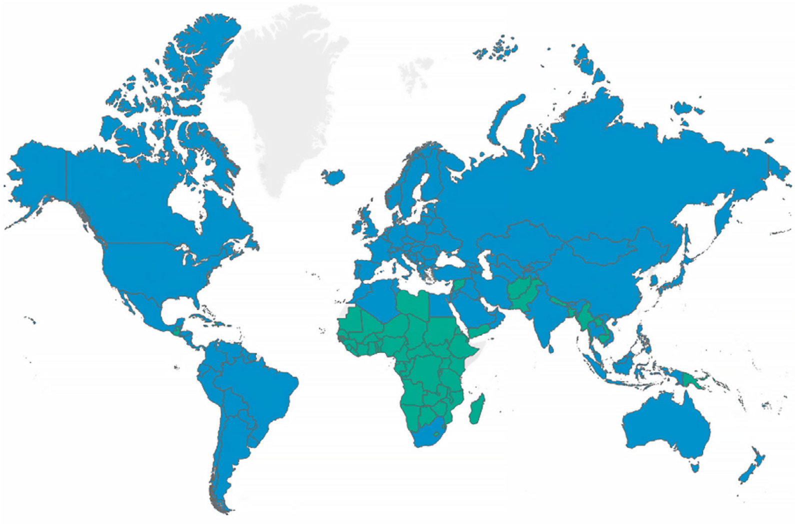


Fig A-3: Barplot of the 10 components (over 60% explained variance) and the respective eigenvalues of the 3 indicators with more weight for the global PCA analysis. A short description of each indicator is also presented (bottom).

1. Total net official development assistance to medical research and basic health sectors per capita (US$), by recipient country
2. Cheapest brand of cigarettes - price in currency reported
3. Road length
4. Number of vets
5. Risk communication
6. Violent Conflict probability
7. Offering help to quit tobacco use
8. Compliance with international health regulations
9. Prevalence of anaemia in children under 5 years (%)
10. Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol), three-year average with 95%CI
11. Alcohol, unrecorded per capita (15+) consumption (in litres of pure alcohol)
12. Population living in urban areas (%)
13. Access to Cities
14. New cases tested for RR-/MDR-TB (%)
15. Tuberculosis treatment coverage
16. Most sold brand of cigarettes - price in currency reported
17. Out-of-pocket expenditure (OOP) per capita in US$
18. Pneumoccocal conjugate vaccines (PCV3) immunization coverage among 1-year-olds (%)

K=2



K=3

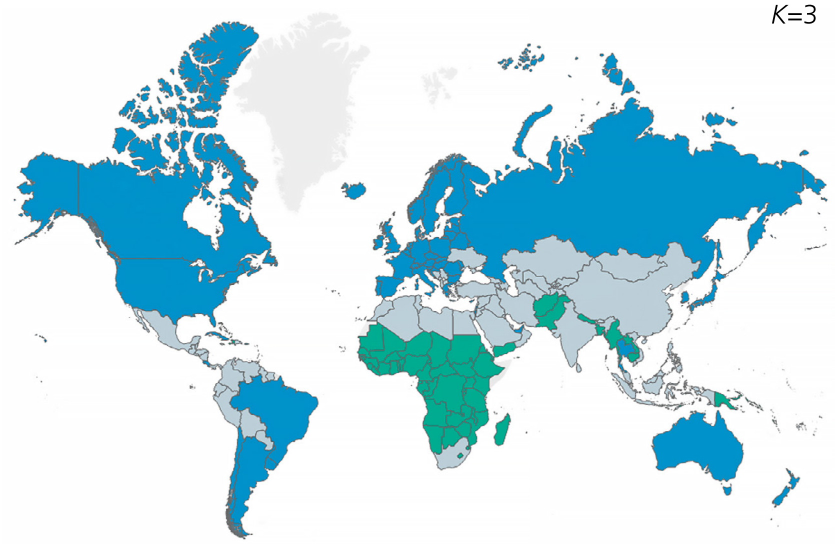


Fig A-4: Worldmap representation of the 2 (top) and 3 (bottom) clusters of the 184 countries of the global analysis.

A.2. Correlation matrix for all countries with mortality data available

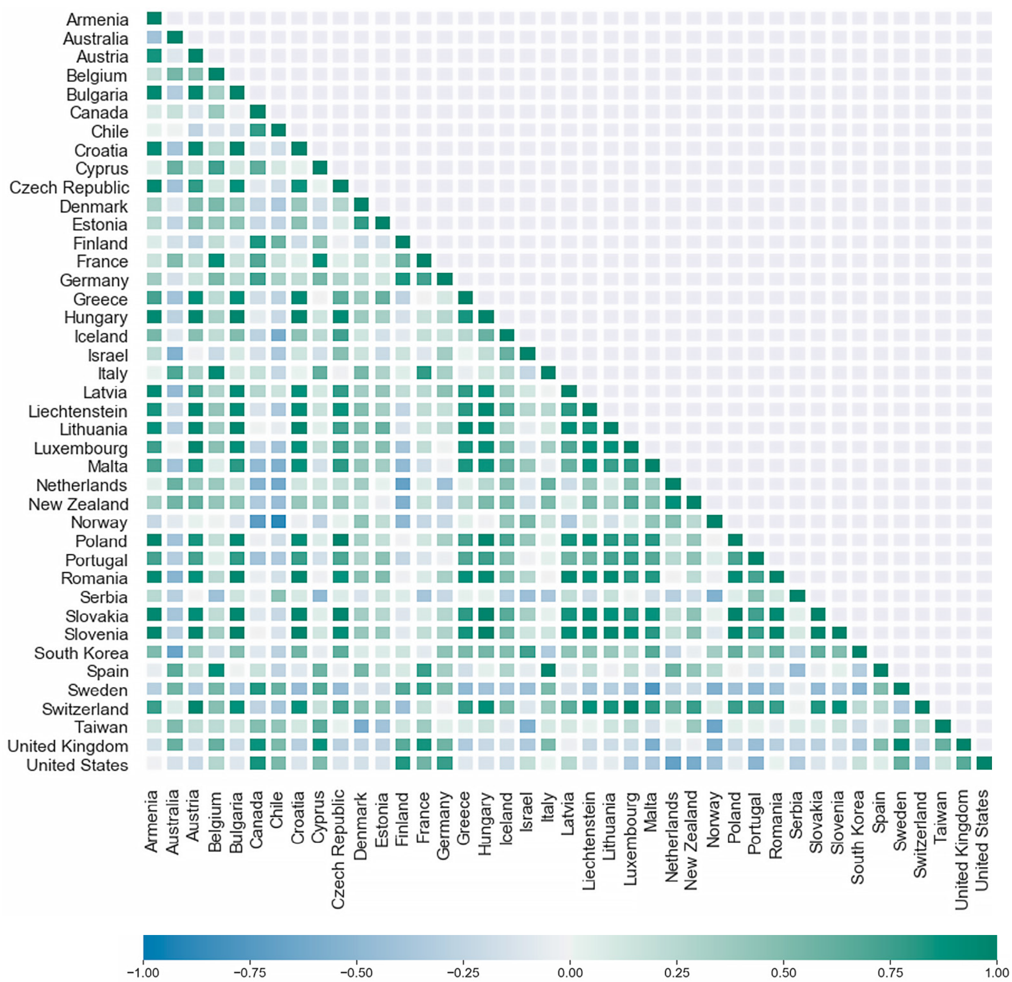


Fig A-5: Monthly series of P-score excess mortality correlation between all countries with mortality data available.

Contacts

Porto – Headquarters
Rua Alfredo Allen 455/461
4200-135 Porto, PORTUGAL

Lisbon – Branch Office
Av. Prof. Gama Pinto 2
1649-003 Lisboa, PORTUGAL

Phone: (+351) 220 430 300
Email: info@fraunhofer.pt
Website: www.fraunhofer.pt