Taking the fight to COVID-19

Understanding how the UK's response to COVID-19 compares with that of other countries and why

V2 Edit

Contents

1. Analytical approach and rationale

In our analysis, we have answered the brief and have gone beyond it, in an engaging way

I AM SMARTICUS

The Brief

An analysis of the global COVID-19 data, compared with and contrasted to the UK data

- A good overview of the data, showing the key issues around the global pandemic
- Commentary on the information
- Tease out and answer questions that arise
- Give the user the ability to interact with the data

We decided to go beyond the brief

The key drivers/predicters of success in fighting the pandemic

- Which countries did better and why?
- How could we segment countries, based on their success in fighting the pandemic

Before we started our analysis, we checked and filtered the world data for consistency

1. Countries removed

Non-country locations (e.g. continents) were removed as it was not known how this data was calculated.

Countries without sufficient weekly information were removed, i.e. Cayman Islands, Comoros, Hong Kong, Lesotho, Marshall Islands, Samoa, Solomon Islands, Tajikistan, Vanuatu and Yemen)

Countries with data limited to late 2020 or early 2021 were also removed, i.e. Anguilla, Bermuda, Faeroe Islands, Falkland Islands, Gibraltar, Guernsey, Isle of Man, Jersey, Macao, Micronesia, Northern Cyprus, Saint Helena, Turks and Caicos Islands and Vanuatu.

2. Consistent time period

The majority of countries reported data from week 15 of 2020 (w/c 06/04/2020) and week 8 of 2021 (w/e 21/02/2021).

We decided to focus on analysis of the World data on this time period, which was a total of 46 weeks.

3. Variables removed

We decided to concentrate on new cases and new deaths as the most meaningful variables to track COVID-19 over time.

We also decided to look at the data week by week, rather than day by day, as public holidays and country differences meant their would be inconsistencies in daily comparisons. This decision made the ‘new cases smoothed’ and ‘new deaths smoothed’ data unnecessary.

Any data that was not consistently provided by the majority of countries was also removed, e.g. ICU/hospital admissions and patients, new and total tests, vaccination rates, etc.

We filled some data gaps in the world data through regression interpolation, created consistency in the stringency index and removed outliers

4. Regression Interpolation

Where data points in new cases and new deaths were missing on a daily basis, the difference with cumulative totals were used to fill in the gaps.

In Kosovo, life expectancy data was added from: https://data.worldbank.org/indicator/SP.DYN.LE00.IN?locations=XK

5. Stringency Index

In the original data this was provided daily but only updated weekly. This meant that there were often missing values mid-week.

We decided to replace this data with the maximum weekly value, which provided more consistency.

Within the analysis, these were collapsed into

a) average stringency index over whole period

b) maximum stringency index over whole period.

In our key driver analysis, only (a) was found to be significant in the models.

6. Outliers removed

Countries were removed from the analysis if they

a) were below the 2.5th percentile in population size, and/or

b) were below the 2.5th percentile in terms of death rate, and/or

c) were below the 2.5th percentile in terms of case rate.

Countries removed for one or more of these reasons were: Cambodia, China, Dominica, Laos, Liechtenstein, Monaco, Saint Kitts & Nevis, San Marino, Tanzania, Timar, Vatican and Vietnam.

This and removals at Box 1 left us with a data set of 169 countries

As we proceeded with our analysis, we followed two key principles

Focus on Case / Death Rate, not total cases and deaths

In our analysis of the World data, we have focused on cases and deaths per 1 million or per 1,000 people, respectively. This seeks to go beyond the obvious conclusion that more populous countries and regions have more cases and deaths, neutralising the population effect.

Clarity and Consistency

Convert raw data into deciles for consistency and clarity

The distribution of the case rate, death rate and all potential predictors was skewed in the World Data, which was not surprising given the diversity of the countries. Rather than transforming the variables we decided to rank the countries on each measure and group these ranks into deciles.

The regression models for the World data are built entirely from these new decile variables, eliminating extreme outliers and producing better fitting models.

Data Sources - Core

Data Set	Description	Source	Core / Supplementary
UK COVid-19 new cases	Up to 21 February 2021	https://www.bhbia.org.uk/assets/Downloads/specimendate_agedemographic-unstacked-uk.csv Taken from: https://coronavirus.data.gov.uk/downloads/demographic/cases/specimenDate_ageDemographic-unstacked.csv	Core
Rest of World COV-19 data	Up to 26 February 2021	https://www.bhbia.org.uk/assets/Downloads/owid-covid-data.csv Taken from Coronavirus Source Data - Our World in Data	Core

Our submission consists of the following 3 documents

PowerPoint Deck

Providing a walk through the data analyses conducted and commentary on the information.

This deck also included a link to our online dashboard, "Taking the Fight to COVID-19"

Excel File

Drivers of Deciles of Case Rate (cases per million collapsed into 10 equal intervals)

Base size: 169

K=3 Model

R-sq	In-sample	56.9%
	CV	51.6%

Order	Quest	Description	Unstandardised Effect	Correlation		Pratts Measure	Johnson Importance	Johnson Importance adjusted for missing values
4	DEC_human_development_index_mean	Composite index measuring average achievement in: long and healthy life, knowledge and decent standard of living	0.3823	0.6785		45%	36%	34%
2	DEC_median_age_mean	Median age of the population, UN projection for 2020	0.2233	0.6585		26%	32%	30%
3	DEC_female_smokers_mean	Share of women who smoke, most recent year available	0.2117	0.4961		16%	19%	23%
1	DEC_stringency_index_mean	Government Response Stringency Index - mean over period	0.2986	0.2560		13%	13%	13%
						100%	100%	100%

Providing a combination of raw data, advanced analytics and pivot tables so you can see our calculations and interact with the data

Video Animation

Watch some of our BHBIA and BOBI committee members battle it out in the Colosseum! Whose region has been the most successful in their fight against COVID-19 and whose has failed?

2. Executive Summary

Executive Summary – World Data

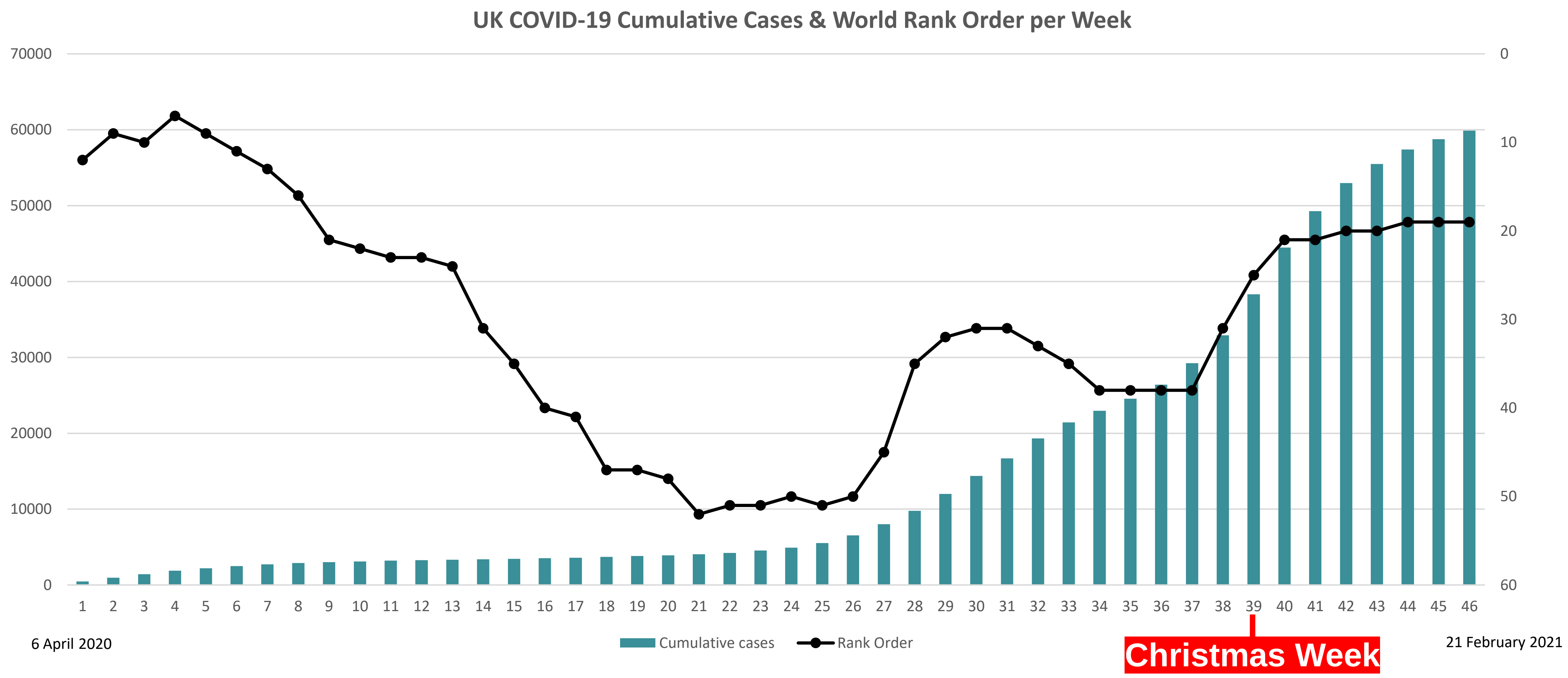
- The UK has among the highest number of cases per million in the World, during the Winter months, but drops out of the Top 50 during the Summer. However, the UK is one of the highest ranked countries in terms of deaths per million throughout the year and never falls below 13th even during the Summer.
- When we consider the key drivers of cases or deaths per million, it is easy to see why the UK rates are so high. The UK is a developed country, with a high median age and a high percentage of women who smoke. The UK government has also imposed strict anti-COVID-19 measures, and these characteristics are all key drivers of high case rate.
  - In considering stringency as a driver, it is important not to assume causation. Nevertheless, the period analysed covered 2 waves of COVID-19 and, if stringency was having the desired effect, a reduction in case rates would be expected during the second wave. This has not been the case.
  - Furthermore, a sharp increase in cases was observed in the run up to Christmas and this coincides with the UK Government's decision to halt the national lockdown on 1st December and move to a three-tier system. A sharp increase in deaths resulted from this rise in cases.
  - The UK has a high deaths-to-cases ratio and an analysis of the drivers of deaths per 1000 reported infections sheds some light on why this might be. The most important driver of deaths per 1000 reported infections is population size. Many of the more populous countries in the World, such as the USA and Western European countries, have more people in hospitals and care homes, who they failed to protect during the pandemic.
- If we go one step further and segment countries based on the key drivers of case rate, the UK belongs to the segment most likely to have a high case rate, i.e. that with a high median age and high proportion of females who smoke. Other countries that fall into this segment include the US, France and Germany, but also Australia and New Zealand who's caseloads have been much lower than the UK's.

3. How the UK measures up

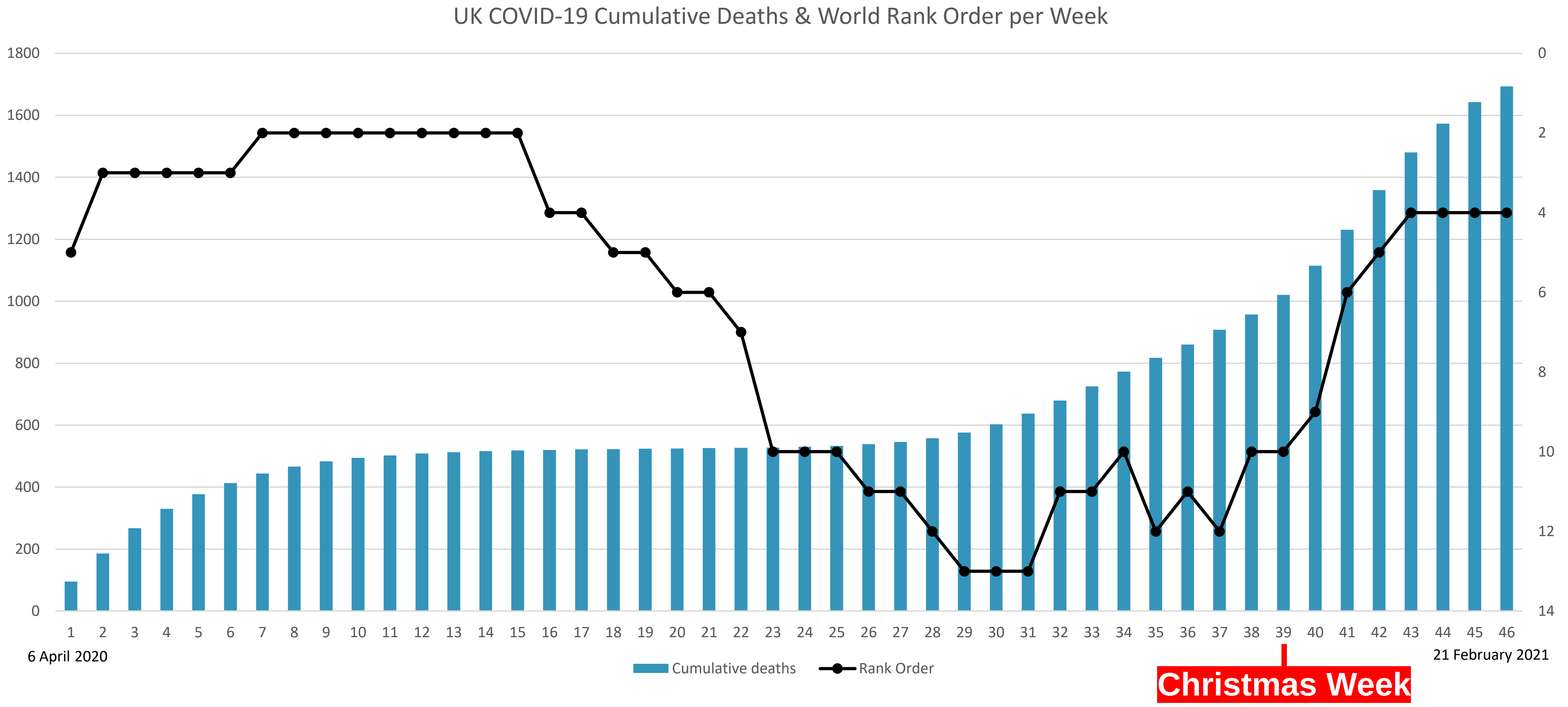
Case rates and death rates by country

- Our Excel file includes the following datasets and analyses:
    - Raw data used for our World and England data analyses, after our checks and data cleaning were complete.
    - Graphs shown in the following two slides which show how the COVID-19 pandemic has evolved over time in the UK and where the UK ranks on case rate and death rate internationally.
- In addition, we have created a bar chart race to show how the case rates and death rates have developed over time. Please follow this link and select the Rest of the World option.
    - These bar chart races can be filtered by continent which enables the viewer to compare and contrast geographically similar countries
- In the UK, the media would have us believe that we have failed in our fight against COVID-19. While we certainly could have done better, it is clear that our cases and deaths per million are seasonal. We are among the worst in the Winter months, but drop in the rankings in the Summer.

The UK has among the highest case rates in the Winter, but drops down the rankings in the Summer. September, when children returned to school and December, in the run up to Christmas, saw sharp increase in cases.



The UK is also among the worst countries in the world in the Winter months on death rate and never drops out of the Top 13 on this metric. The increase in cases in September and December translated into a sharp rise in deaths.



4. The drivers of case and death rates internationally

World Data Analysis

- Please refer to our Excel file once again, now from the tab titled ‘World Analysis:
  - CCR-Johnson’s key driver analysis on case rates, death rates and deaths per 1000 infections
  - Correlation analysis to assess whether drivers operate independently from each other
  - CHAID segemtnation analysis
- In addition, we have created an interactive dashboard, so you can see how countries compare with each other on case rates and death rates, either at a total level or week by week, along with the drivers of case and death rate. Please follow this link and select the Rest of the World option.

The key drivers methodology that we used combines Correlated Component Regression (CCR) and in important index called Johnson’s Relative Weights

A high dimensional regression algorithm known as Correlated Component Regression was used which delivers models which are maximally robust when number of observations are small and correlations between predictors are large as was the case in the Covid data sets.

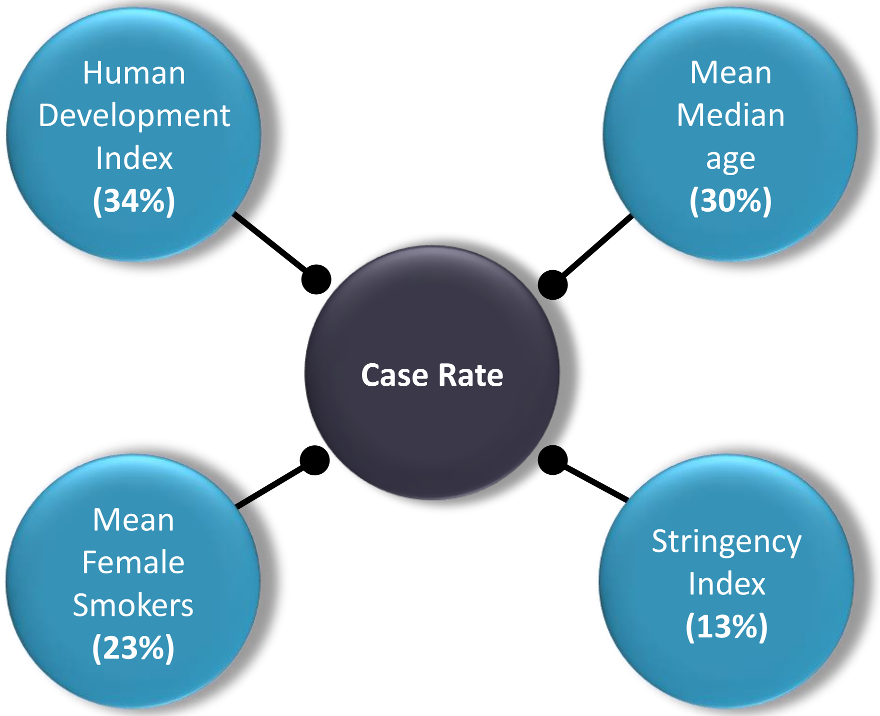
- A regression model approach which delivers a Shapley-Value-like index, for as many predictors as we need, that works for extreme situations: Small samples, many highly correlated predictors.
- Works within all common types of modelling framework: Logistic and ordinal, as well as linear models.
- Can deliver, if needed, two sets of results: One with optimal variables screening, and one without.
- Uses Cross-validation at core to select most stable model specification and optimal number of predictors if screening.

This uses a combination of our Correlated Component Regression (CCR) methodology and an Importance Index being widely adopted called Johnson’s Relative Weights. This method allows an almost 100% approximation of Shapley-Value coefficients, but with many advantages over the Shapley-Value method:

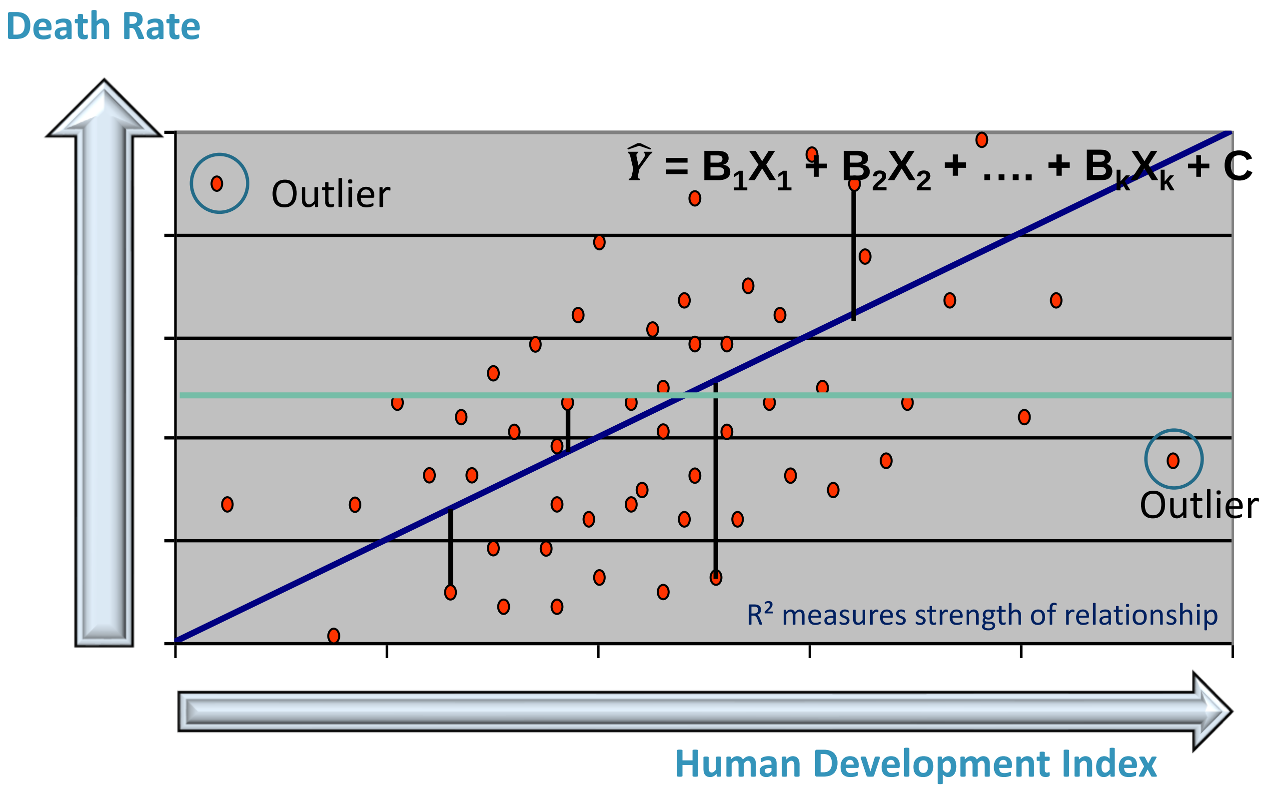
- Quicker to estimate for large data sets, so faster turnaround.
- No limit on the number of predictors (Shapley becomes cumbersome after about 10)
- Can be applied to Logistic and Ordinal as well as Linear Regression models
- Unlike some other importance measurement methods (e.g. Pratts) doesn’t result in negative importance for Suppressor variables (where effect size is opposite sign to correlation).

Our driver analysis looks at both the Johnson Importance, adjusted for missing values, as well as the unstandardised effect of each driver

Johnson's Importance, adjusted for missing values



Unstandardised Effect (Linear regression attempts to fit a line to estimate Y as a function of X)



Key Drivers – Case Rate
One of the key drivers of cases is the stringency index, which suggests that lockdowns are not having the desired effect.

Quest	Description	Effect	Johnson Importance adjusted for missing values
DEC_human_development_index_mean	Composite index measuring average achievement in: long and healthy life, knowledge and decent standard of living	0.3823	34%
DEC_median_age_mean	Median age of the population, UN projection for 2020	0.2233	30%
DEC_female_smokers_mean	Share of women who smoke, most recent year available	0.2117	23%
DEC_stringency_index_mean	Government Response Stringency Index - mean over period	0.2986	13%

Crossvalidated r-squared = 52%

When assessing the key drivers of case rates, it is important to consider both the CCR-Johnson's importance as well as the standardised effect.

The most important driver of cases is the human development index. This attribute also has the greatest effect on the number of cases. The more developed a country is, the more cases it reports. This may be because more developed countries have the infrastructure to detect cases better and report more accurate data. For every decile increase in the human development index, a country will increase its case rate by 0.3823 deciles.

The median age in a country is the second most important driver of cases, although it has a smaller unstandardized effect. We know that COVID-19 tends to impact older people more and it is logical that cases in the elderly are more likely to be diagnosed and reported.

The share of women who smoke also contributes to the number of cases in a country, although once more the magnitude of the effect is smaller. Again, logic would suggest that COVID-19 cases in smokers are more likely to be detected.

The more stringent a government's response to the pandemic, the more cases it is likely to have. Some might suggest that more developed countries have put stringent anti-COVID measures in place, but the data does not support this, with weak correlation between the human development index and the stringency index.

It is logical that government with a high volume of cases are more likely to impose a stricter lockdown. However, it does not then follow that stricter lockdowns reduce the number of cases effectively.

The implication is that stringent lockdowns are not effective.

Key Drivers – Death Rate
One of the key drivers of death is the stringency index, which suggests that lockdowns are not having the desired effect.

The drivers of death rate are somewhat different to those of case rate.

The most important driver of death rate is the share of the population aged over 65 years. Countries with older populations are likely to suffer from higher COVID death rates than those with younger populations. In fact, for every decile increase on this factor, a country will move up 0.2962 deciles on death rate.

As was the case with case rate, the proportion of females who smoke is also a strong predictor of death rate. This feature is closely correlated with the human development index but is also a strong predictor in its own right.

Once again, the human development index is a strong driver. Again, our hypothesis is that more developed countries report deaths related to COVID-19 more consistently than less developed countries.

The stringency index is also a key driver of death rate. In this case, it also has the largest standardized effect on death rate, of all the drivers. If prevention of death is the primary objective of lockdowns, this evidence suggests that they are not meeting this goal.

Quest	Description	Effect	Johnson Importance adjusted for missing values
DEC_aged_65_older_mean	Share of the population that is 65 years and older, most recent year available	0.2962	29%
DEC_female_smokers_mean	Share of women who smoke, most recent year available	0.2595	28%
DEC_human_development_index_mean	Composite index measuring average achievement in: long and healthy life, knowledge and decent standard of living	0.226	26%
DEC_stringency_index_mean	Government Response Stringency Index - mean over period	0.3263	17%

Crossvalidated r-squared = 47%

Key Drivers – Deaths per 1000 reported infections
The strongest driver of deaths per 1000 reported infections is population, perhaps because larger, developed countries have more people vulnerable to infection in hospitals and homes

Quest	Description	Effect	Johnson Importance adjusted for missing values
DEC_population_mean	Population in 2020	0.3067	65%
DEC_population_density_mean	Number of people divided by land area, measured in square kilometers, most recent year available	-0.1975	26%
DEC_aged_65_older_mean	Share of the population that is 65 years and older, most recent year available	0.1317	9%

Crossvalidated r-squared = 9%

We also looked and the key drivers of deaths per 1000 reported infections within the World data. Although this model is weaker than the other two driver analyses, with a CV r-squared below 10%, it nevertheless reveals some interesting insights and neutralizes the reporting effect whereby developed markets tend to have more developed reporting processes

The most important driver of deaths per 1000 reported infections is population size. This may be due to a variety of factors.

Firstly, several populous countries such as India and Indonesia have poor standards of healthcare and poor COVID detection rates. COVID-sufferers are often only detected upon hospitalization and such patients more often succumb to the disease than out-patients.

There are also several populous countries with well developed healthcare provision, such as the USA and the UK. Such countries have more hospitals per capita and there is growing evidence that in such countries a high proportion of infections are picked up in hospitals. Such patients tend to be vulnerable and more likely to die of COVID-19. The same is true of the number of care homes in such countries, where the death per case rate has been high.

For every decile increase in population size, a country's deaths per 1000 reported infections increase by 0.3067 of a decile.

The second most important driver of deaths per 1000 reported infections is population density. In fact, less densely populated countries have higher deaths per 1000 reported infections .

In densely populated countries, people are more likely to encounter each other, driving the caseload up. These cases are also more likely to be reported. However, due to high standards of healthcare, the deaths per 1000 reported infections in such countries tends to be lower.

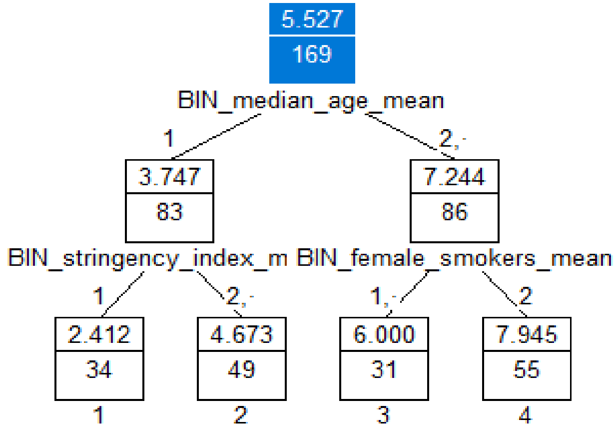
It is not surprising that the third driver of deaths per 1000 reported infections is the percentage of the population that is over 65 years old. The only surprise, perhaps, is that this is not a stronger driver, but it would seem that the types of people being infected and the nature of the reported cases, i.e. hospitalization or not, are more important facets.

We then conducted CHAID analysis, using the key drivers of case rate and death rate, to segment countries

I AM SMARTICUS

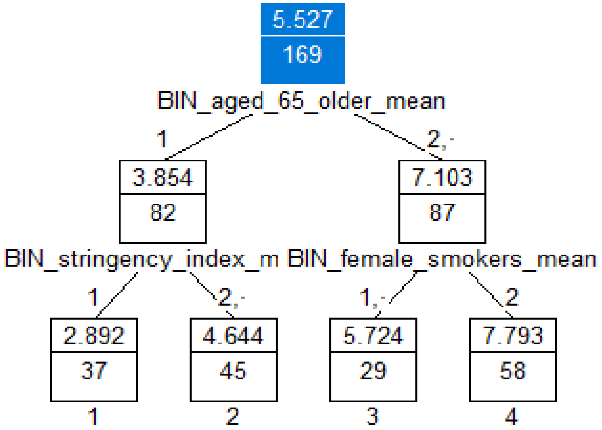
Case rate

Segment	Base	% Age Median Age	Stringency index	% Female smokers	Mean Decile (death rate) Unweighted	Mean Decile (death rate) weighted by population size
1	34	Low	Low		2.41	2.21
2	49	Low	High		4.67	4.42
3	31	High		Low	6.00	4.89
4	55	High		High	7.95	8.11



Death rate

Segment	Base	% Age 65+	Stringency index	% Female smokers	Mean Decile (death rate) Unweighted	Mean Decile (death rate) weighted by population size
1	37	Low	Low		2.89	3.16
2	45	Low	High		4.64	4.53
3	29	High		Low	5.72	5.01
4	58	High		High	7.79	8.36



Case Rate Segmentation
The UK is in the segment most likely to have a high case rate

Young (low median age)

Carefree (low stringency)

Mean decile case rate 2.21

e.g. Fiji, Afghanistan

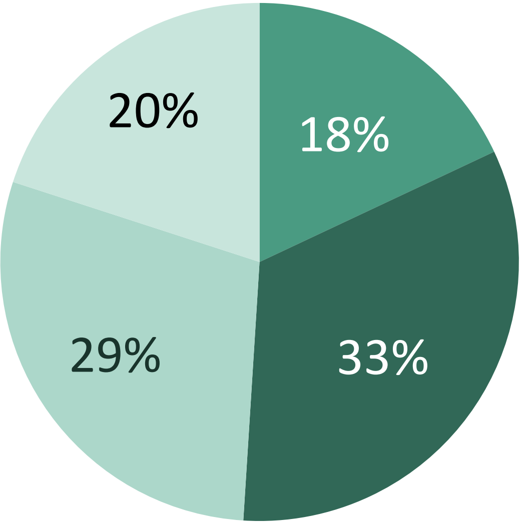
Old (high median age)

Smoke-free (low fem smoke)

Mean decile case rate 4.89

e.g. Barbados, Taiwan

All countries



Base: n=169

Young (low median age)

Strict (high stringency)

Mean decile case rate 4.42

e.g. Algeria, Bangladesh

Old (high median age)

Smoky (high fem smoke)

Mean decile case rate 8.11

e.g. UK, Brazil

Death Rate Segmentation - a similar results, with slightly different drivers. The UK once again belongs to the segment most likely to have a high death rate

Young (low % >65)

Carefree (low stringency)

Mean decile death rate 3.16

e.g. Fiji, Afghanistan

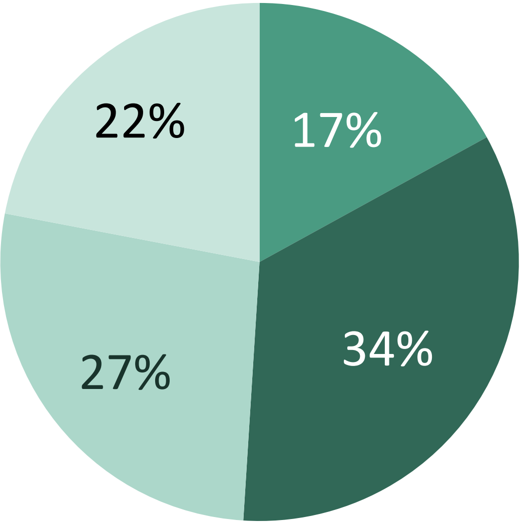
Old (high % >65)

Smoke-free (low fem smoke)

Mean decile death rate 5.01

e.g. Barbados, Taiwan

All countries



Base: n=169

Young (low % >65)

Strict (high stringency)

Mean decile death rate 4.53

e.g. Algeria, Bangladesh

Old (high % >65)

Smoky (high fem smoke)

Mean decile death rate 8.36

e.g. UK, Brazil

Thank you