I AM SMARTICUS

# Taking the fight to COVID-19

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

# Contents

# 1. Analytical approach and rationale

I AM SMARTICUS

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

## 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
- Which regions within England did better and why?

# 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

# In England, we identified supplementary data to enrich our key driver and segmentation analysis

## 1. Supplementary Data

Additional data from external sources (see Data Sources - Supplementary) was matched to the core data by Lower Tier Local Authority (LTLA). The core English Covid dataset contained 315 of the 317 LTLAs on the Annual Population Survey/Labour Force Survey dataset*

*City of London and Hackney were merged on the Covid data, Isles of Scilly were merged with Cornwall

## 2. LTLA Summaries Created

The data matched was created from data at lower levels of geography (Lower Super Output Areas = LSOAs), which were averaged in two alternative ways to create the LTLA (Lower Tier Local Authority) summaries:

a) rank average of the LSOA rank for each score (used to create deciles for each measure) and

b) average raw scores across an LTLA's constituent LSOAs

## 3. Consistent Time Period

A period of 48 weeks was chosen where there was consistent data on deaths and cases from across 314 LTLAs. The date range was from Calendar week 13 2020 (w/c 23/03/2020) to Calendar week 8 2021 (w/e 21/02/2021).

Due to missing data at the start of the series, North Devon was dropped from the analysis

# In England, we also filled data gaps through regression interpolation before running final data checks

I AM SMARTICUS

## 4. Regression Interpolation

Four of the 314 regions did not have matchable data for the proportion aged 16+ in employment and the proportion of non-white ethnicity due to structural changes in the definitions of LTLAs during 2020. The effected LTLAs were Aylesbury Vale, Chiltern, South Bucks and Wycombe. The missing values on these two variables were interpolated for these four LTLAs using the regression interpolation method. This ensured that the English data set of 314 LTLAs was complete and that no values remained missing.

## 5. Data Checks

Exploratory analysis of death rate, case rate and all average LTLA predictor scores (average raw scores across an LTLA's constituent LSOAs) suggested that distributions were approximately normally in most cases and not overly skewed. A decision was made to work with these averaged raw scores for developing regression models

# 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 both the World and England 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.

The England modelling was based on raw data, but then reported as deciles for consistency and clarity.

# 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<br>Taken from:<br>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<br>Taken from Coronavirus Source Data - Our World in Data | Core |

# Data Sources - Supplementary

| Data Set | Description | Source | Core / Supplementary |
|---|---|---|---|
| English indices of deprivation 2019 | Released: 26 September 2019<br>Data used: File 10: local authority district summaries<br>Variables include:<br>• Index of Multiple Deprivation<br>• Income<br>• Employment<br>• Education, Skills and Training<br>• Health and Disability<br>• Crime<br>• Barriers to Housing and Services<br>• Living Environment<br>• Income Deprivation Affecting Children<br>• Income Deprivation Affecting Older People | https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019 | Supplementary |
| UK COVid-19 new deaths within 28 days of test | Date of Data: 8 March 2021 | https://coronavirus.data.gov.uk/api/v2/data?areaType=ltla&metric=newDeaths28DaysByDeathDate&format=csv | Supplementary |

# Data Sources - Supplementary

| Data Set | Description | Source | Core / Supplementary |
|---|---|---|---|
| Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland | Released: 24 June 2020<br>Data used: MYE2 – Persons, MYE5<br>Date of Data: Mid 2019 – Mid 2020<br>Geography: Local authority: district / unitary (England only)<br>Variables: Age-0-90+, Area | https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland/mid2019april2020localauthoritydistrictcodes/ukmidyearestimates20192020ladcodes.xls | Supplementary |
| Annual Population Survey / Labour Force Survey | Released: 26 Jan 2021<br>Geography: local authority: district / unitary (England only)<br>Date of Data: Sept 2020<br>Variables selected:<br>T18:1 (All Ages - All : All People )<br>T18:4 (All Ages - White: All People)<br>T01:1 (All aged 16 & over - All : All People)<br>T01:7 (All aged 16 & over - In employment : All People)<br>T01:16 (All aged 16 & over - Unemployed : All People)<br>T01:19 (All aged 16 & over - Inactive : All People) | https://www.nomisweb.co.uk/ | Supplementary |

# 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.

# Executive Summary – England Data

- Ethnicity is one of the most important drivers of COVID cases in England and also has the biggest effect on a region's case rate decile. It is also a Top 5 driver of death rate and has the largest effect on a region's death rate decile.
  - Although ethnicity is correlated to a moderate level with other important drivers, such as crime rates, income and education, it clearly is a strong driver in its own right. In England, it would seem that we have not done enough to protect BAME communities from COVID infection and death.
- Crime, low income and poor education are also important drivers of COVID cases and deaths in England. More deprived areas have certainly been hit by the pandemic more acutely than affluent areas.
- Age is a critical factor in determining case rates and death rates in England. Those regions with high proportions of their population aged 70+ tend to have lower case rates, but higher death rates. This underlines the dangerous nature of the disease in the older generation.
- The role that age plays can be seen when we segment regions based on case rates and death rates
  - Those regions with high crime rates and a low proportion of over 70s are most likely to experience high COVID-19 case rates, for example Ealing and Liverpool.
  - However, if we segment regions based on their deaths per 1,000 people, high proportion over 70 drives death rate, though this is a less of a factor overall. It is those regions with poor education and high crime rates that are most at risk.

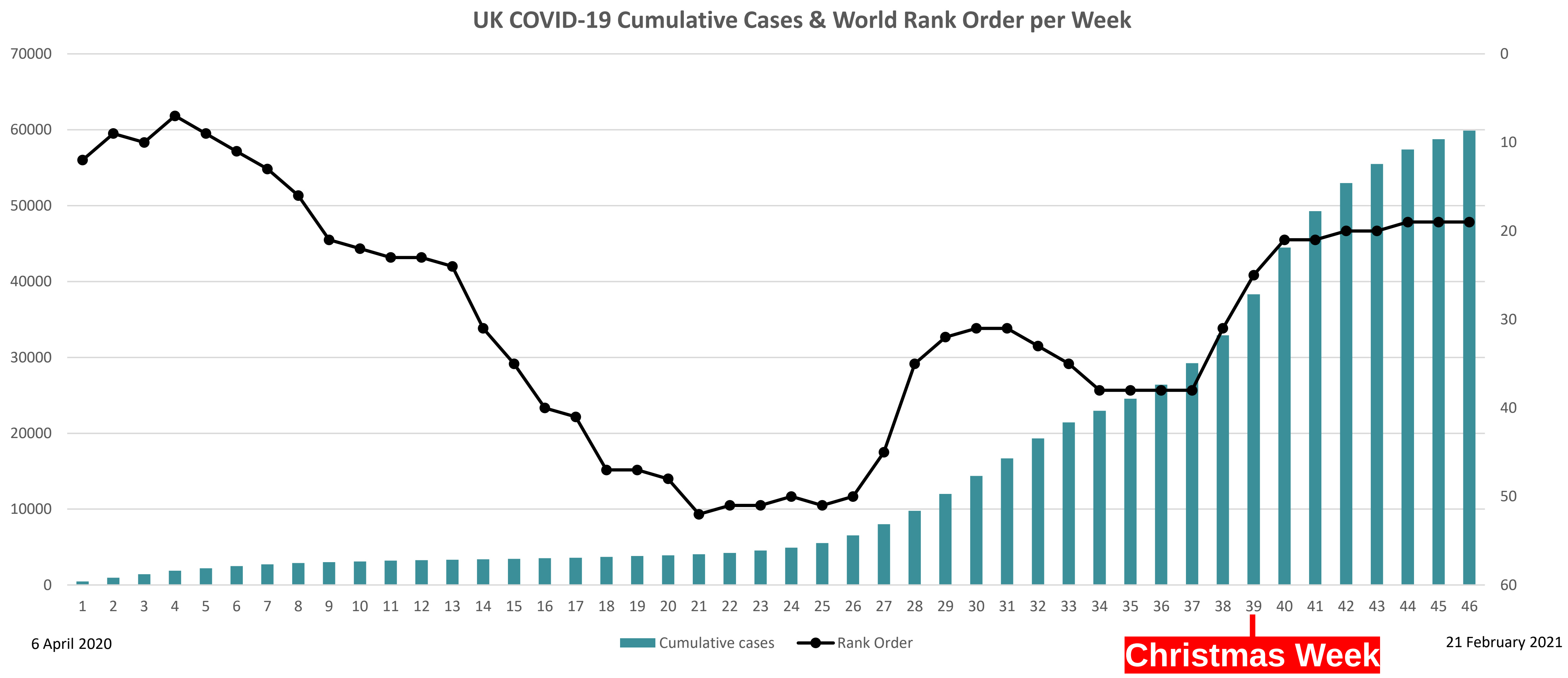
# 3. How the UK measures up

I AM SMARTICUS

# Case rates and death rates by country

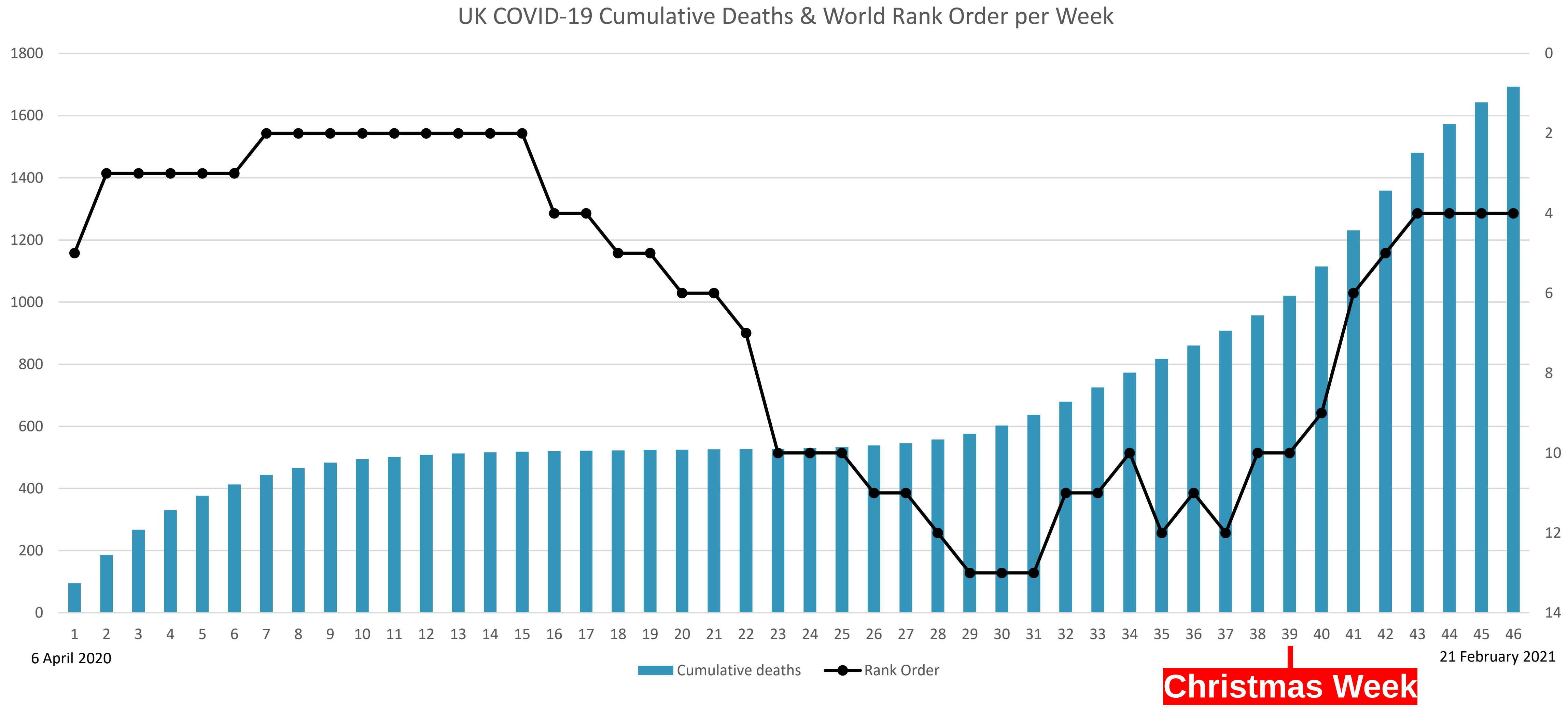
- Our Excel file includes the following datasets and analyses:
    - Raw data used for our World and England data analyses, after our checks and data cleaning were complete.
    - Graphs shown in the following two slides which show how the COVID-19 pandemic has evolved over time in the UK and where the 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.
    - 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.

I AM SMARTICUS



# 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 segmentation 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.

# 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.

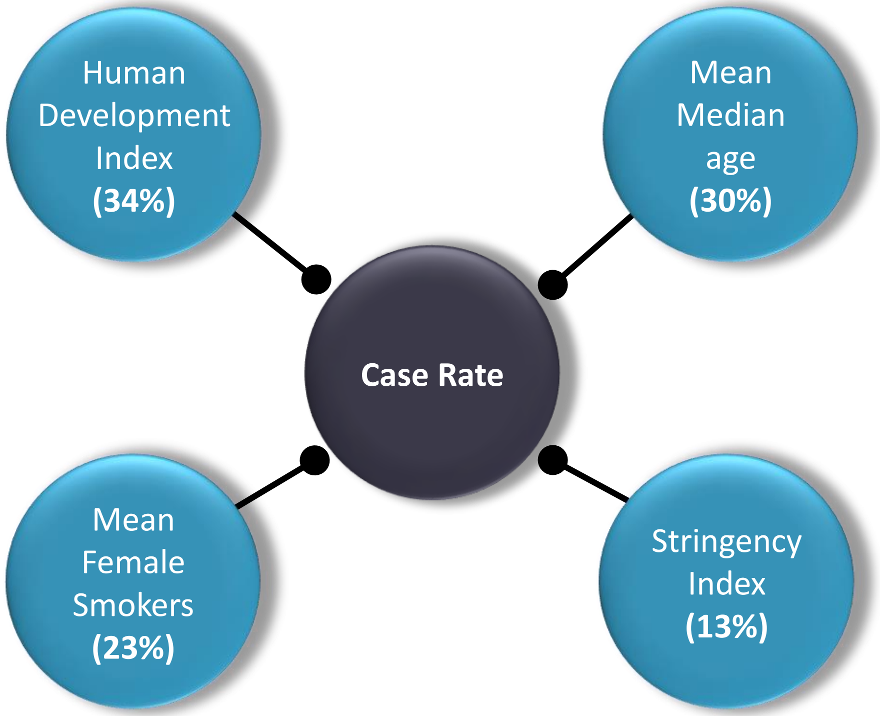
- A regression model approach which delivers a Shapley-Value-like index, for as many predictors as we need, that works for extreme situations: Small samples, many highly correlated predictors.
- Works within all common types of modelling framework: Logistic and ordinal, as well as linear models.
- Can 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:

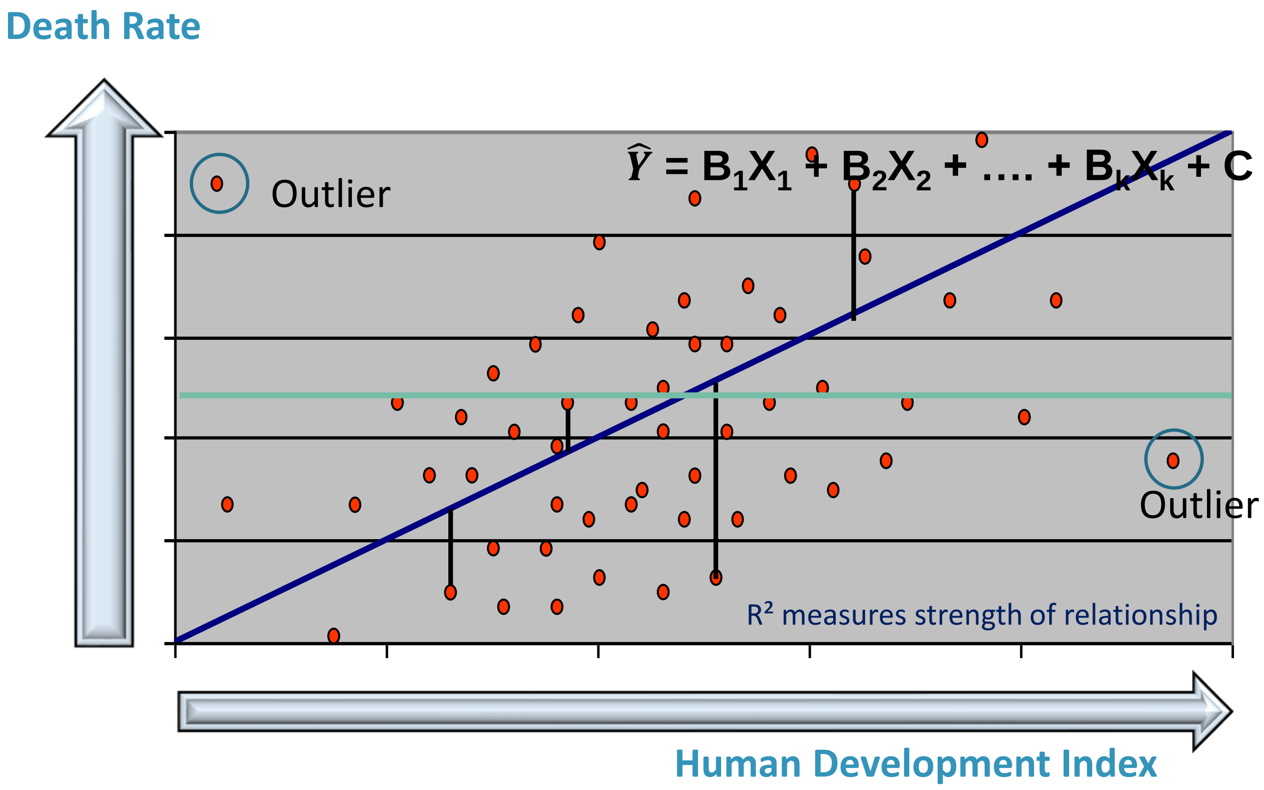
- Quicker to estimate for large data sets, so faster turnaround.
- No limit on the number of predictors (Shapley becomes cumbersome after about 10)
- Can be applied to Logistic and Ordinal as well as Linear Regression models
- Unlike some other importance measurement methods (e.g. Pratts) doesn't result 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 .

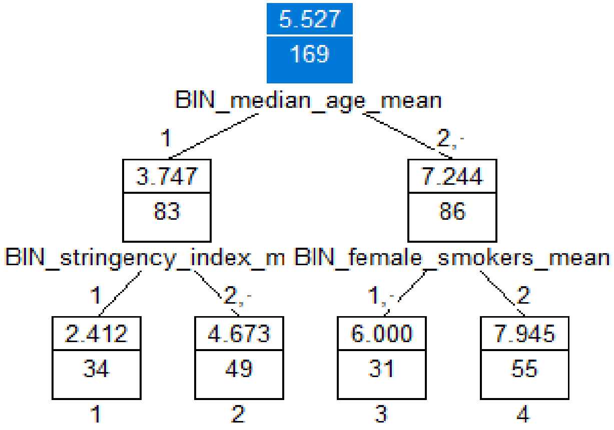
In densely populated countries, people are more likely to encounter each other, driving the caseload up. These cases are also more likely to be reported. However, due to high standards of healthcare, the deaths per 1000 reported infections  in such countries tends to be lower.

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

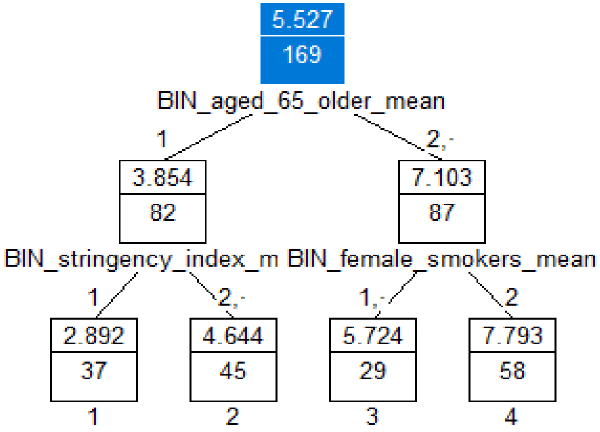
## Case rate

| 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

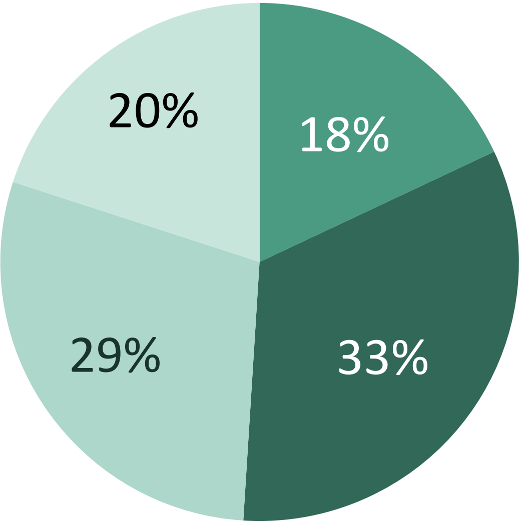
Old (high median age)

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

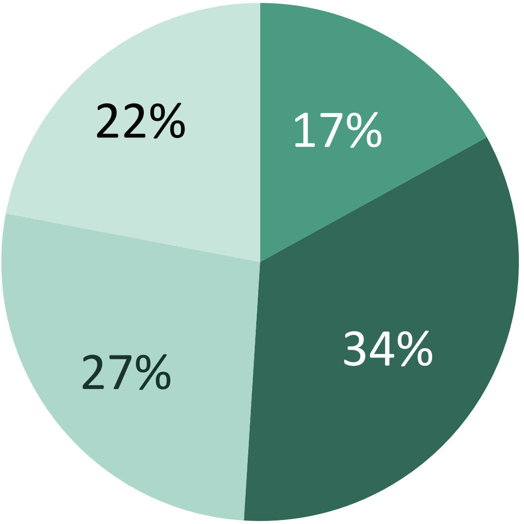
Young (low % >65)

Carefree (low stringency)

Mean decile death rate 3.16

e.g. Fiji, Afghanistan

**All countries**




Base: n=169

Old (high % >65)

Smoke-free (low fem smoke)

Mean decile death rate 5.01

e.g. Barbados, Taiwan

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

# 5. The drivers of case and death rates in England

# England Data Analysis

- Please refer to our Excel file once again, now from the tab titled ‘England Analysis:
  - CCR-Johnson’s key driver analysis on case rates, death rates and deaths per 1000 infections.
    - For the England data, we re-ran this by age brand to assess the impact of age more thoroughly.
  - Correlation analysis to assess whether drivers operate independently from each other
  - CHAID segmentation analysis
- We have also created a heat map for the England data, as well as an interactive dashboard, so you can see how regions 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.

# Key Drivers – Case Rate
## Ethnicity is one of the most important drivers of case rate in England and has the biggest effect on a region's case rate.

| Quest | Description | Effect | Johnson Importance |
|---|---|---|---|
| Crime_score | Risk of personal and material victimisation at local level | 0.2665 | 21% |
| Ethnicity_NonWhite | % Non-White ethnicity background | 0.5408 | 21% |
| Prop_Age_70plus | % of Population - Aged 70+ | -0.2058 | 18% |
| Income_score | % in population experiencing deprivation relating to low income | 0.2474 | 15% |
| Eduk_score | Lack of attainment and skills in the population | 0.1043 | 10% |
| Population_density_sq.km | Population density per sq km | -0.2651 | 6% |
| All_ages_population | Population - all ages | -0.1246 | 5% |
| Barriers_score | Physical and financial accessibility of housing and local services | -0.1089 | 2% |
| Living_score | Quality of both the 'indoor' and 'outdoor' local environment | -0.1117 | 1% |

There has been some coverage in the UK press that BAME communities are at greater risk of COVID-19 and our analysis support this. In fact, ethnicity is one of the most important drivers of case rate in England and also has the biggest effect on a region's case rate decile.

Crime rates are an equally important driver, although the effect on a region's case rate decile is lower.

The correlation between ethnicity and crime is not high.

The percentage of the population aged 70+ has a negative impact on case rates, with younger populations more likely to leave the home, opening them up to infection.

More deprived regions, in terms of income and education, are more likely to have higher case rates.

Income deprivation is strongly correlated with both crime levels and ethnicity.

The more densely populated a region is, the lower its case rates are likely to be. This may be because people in less densely populated regions have to travel further for work, school or shopping. It may also be because rural locations have less access to home delivery.

Other, weaker drivers of case rates include population size (negatively), deprivation of housing and local services and quality of life scores.

In the Excel file, you will see additional drivers analysis run for case rates in different age groups.

Crime rates, ethnicity and income deprivation consistently emerge as the most important drivers of case rates across age groups. Furthermore, ethnicity has the biggest effect on a region's case rate at all ages.

# Key Drivers – Death Rate
## Crime, poor education and low income are the strongest drivers of death rate. Ethnicity is also an important driver and has the greatest effect on death rate.

I AM SMARTICUS

Crime rates, poor education and low incomes are the most important drivers of death rate in England.

Of these three characteristics, crime has the largest effect on death rate, followed by income deprivation.

Interestingly, income deprivation among the over 60s is a negative driver of death rates, perhaps with poorer elderly people more likely to be staying home.

Although ethnicity is a weaker driver of death rate than it was of case rate, it is nevertheless one of the top five drivers of death rate and has the biggest effect on a region's death rate decile.

Accessibility of housing and local services is a negative driver of death rate.

Age is an important driver of death rate, with those regions with a high proportion of over 70s and a low proportion of under 25s more likely to have a high death rate.

Quality of life, population density and population size all impact death rates negatively.

| Quest | Description | Effect | Johnson Importance |
|---|---|---|---|
| Crime_score | Risk of personal and material victimisation at local level | 0.3714 | 21% |
| Eduk_score | Lack of attainment and skills in the population | 0.1774 | 19% |
| Income_score | % in population experiencing deprivation relating to low income | 0.2916 | 18% |
| Income_older_score | % of those aged 60+ experiencing income deprivation | -0.1218 | 9% |
| Ethnicity_NonWhite | % Non-White ethnicity background | 0.4696 | 7% |
| Barriers_score | Physical and financial accessibility of housing and local services | -0.1570 | 7% |
| Prop_Age_70plus | % of Population - Aged 70+ | 0.2991 | 5% |
| Prop_Age_0_24 | % of Population - Aged 0-24 | -0.2376 | 5% |
| Living_score | Quality of both the 'indoor' and 'outdoor' local environment | -0.2134 | 4% |
| Population_density_sq.km | Population density per sq km | -0.1772 | 3% |
| All_ages_population | Population - all ages | -0.1298 | 2% |

# Key Drivers – Deaths per 1000 reported infections
# The key drivers of deaths per 1000 reported infections are overwhelmingly age related.

| Quest | Description | Standardised Effect | Johnson Importance |
|---|---|---|---|
| Prop_Age_70plus | % of Population - Aged 70+ | 0.6501 | 60% |
| Prop_Age_0_24 | % of Population - Aged 0-24 | -0.1576 | 39% |
| Eduk_score | Lack of attainment and skills in the population | 0.1157 | 1% |

The strongest driver of death per case is the percentage of the population aged 70 years and over. As COVID-19 is more deadly in the over 70s, this finding is logical.

Furthermore, the percentage of the population aged over 70 has a strong effect on the deaths per 1000 reported infections decile.

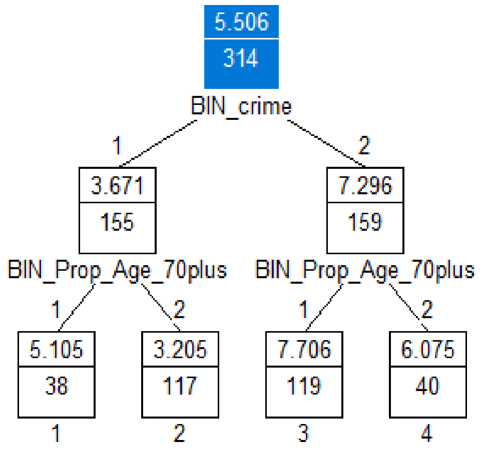
The second most important driver of deaths per 1000 reported infections is the percentage of the population aged under 25. This has a negative effect on the deaths per 1000 reported infections decile. It is the proportion of older, relative to younger members of the population that drives deaths per per 1000 reported infections.

The education deprivation index is a driver of deaths per 1000 reported infections, but much less important than either of the above-mentioned age-related characteristics.

# We again used CHAID analysis, this time to segment regions in England based on the key drivers of case rate and death rate

**Case rate**

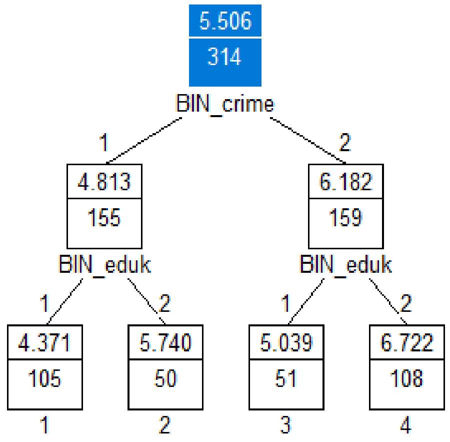
| Segment | Base | Crime deprivation | % of 70+ in popn | | Mean Decile (death rate) Unweighted | Mean Decile (death rate) weighted by population size |
|---|---|---|---|---|---|---|
| 1 | 38 | Low | Low | | 5.11 | 5.18 |
| 2 | 117 | Low | High | | 3.21 | 2.97 |
| 3 | 119 | High | Low | | 7.71 | 7.86 |
| 4 | 40 | High | High | | 6.08 | 6.33 |




**Death rate**

| Segment | Base | Crime deprivation | Education deprivation | | Mean Decile (death rate) Unweighted | Mean Decile (death rate) weighted by population size |
|---|---|---|---|---|---|---|
| 1 | 105 | Low | Low | | 4.37 | 4.31 |
| 2 | 50 | Low | High | | 5.74 | 5.19 |
| 3 | 51 | High | Low | | 5.04 | 4.80 |
| 4 | 108 | High | High | | 6.72 | 6.46 |

5.411252448



# Case Rate Segmentation
## The highest case rates are seen in regions with high crime rates and a low percentage of over 70s, such as Ealing and Liverpool

Low crime rate

High % over 70s

Mean decile case rate 2.97

e.g. S. Northants, Waverley

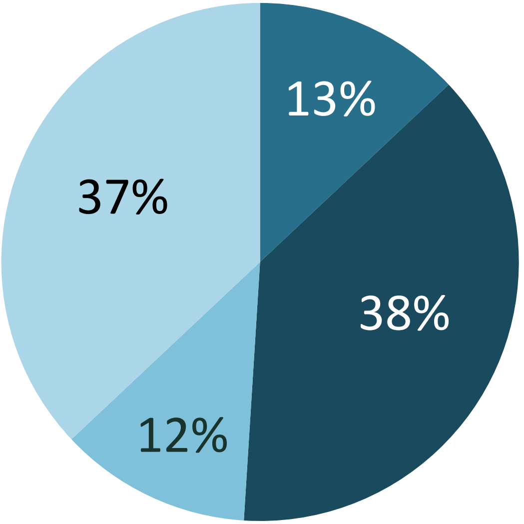
High crime rate

High % over 70s

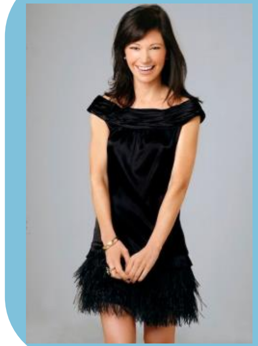
Mean decile case rate 6.33

e.g. Blackpool, Bournemouth

**All regions**




Base: n=314

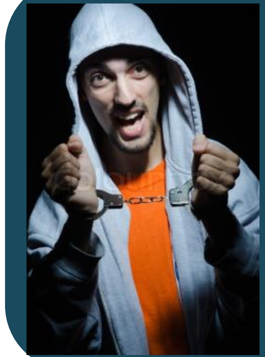


Low crime rate

Low % over 70s

Mean decile case rate 5.18

e.g. Bracknell, Exeter



High crime rate

Low % over 70s

Mean decile case rate 7.86

e.g. Ealing, Liverpool

# Death Rate Segmentation
## Those regions that are deprived due to low education levels and high crime rates are most likely to have high COVID-19 death rates

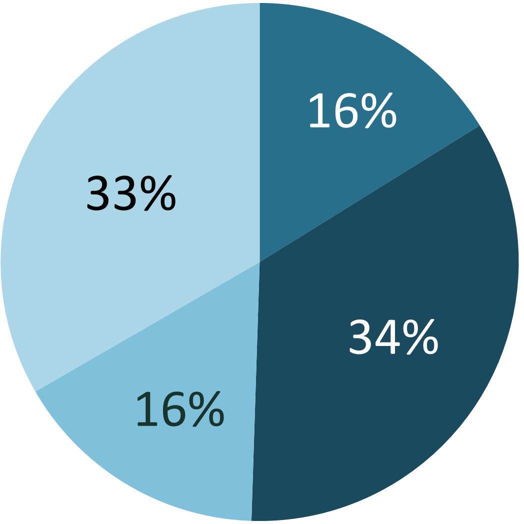
Good education

Low crime rate

Mean decile death rate 4.31

e.g. Reigate and Banstead

**All regions**




Base: n=314

Poor education

Low crime rate

Mean decile death rate 5.19

e.g. Dorset

Good education

High crime rate

Mean decile death rate 4.80

e.g. Westminster


Poor education

High crime rate

Mean decile death rate 6.46

e.g. Gateshead

# 6. Gladiators, are you ready?

# Gladiators: Full programme

| | | |
|---|---|---|
| Aston Le Walls<br>**Northamptonshire** | **v** | Bracknell<br>**Bracknell Forest** |
| Elstead<br>**Waverly** | **v** | Ealing<br>**Ealing** |
| Redhill<br>**Reigate and Banstead** | **v** | Reading<br>**Reading** |
| **Spain** | **v** | **Italy** |

# ROUND 1

*"Despite a higher level of criminal activity, Wilf managed to clinch this round with fewer deaths and cases. Kelly boasts good behaviour but should have worked harder at school to reduce that case rate!"*


*Aston Le Walls*

**South Northamptonshire**

**Cases: 47.25**
*(per 1000 people)*
**Deaths: 1.16**
*(Per 1000 people)*
**Deaths/cases: 2.1%**

**Crime level: 2**
**Low education level: 2**
**Low income level: 1**
**Older age score: 6**

*Bracknell*

**Bracknell Forest**

**Cases: 54.98**
*(per 1000 people)*
**Deaths: 1.27**
*(Per 1000 people)*
**Deaths/cases: 2.3%**

**Crime level: 1**
**Low education level: 5**
**Low income level: 2**
**Older age score: 2**

# ROUND 2

*"Sarah was odds on favourite to lose this round, based on previous crimes committed and a lower level of income, yet somehow Steve was defeated with a higher death rate overall… must be the younger demographic in Ealing!"*

*Elstead*

**Waverley**

**Cases: 44.48**
*(per 1000 people)*
**Deaths: 1.90**
*(Per 1000 people)*
**Deaths/cases: 4.3%**

**Crime level: 2**
**Low education level: 1**
**Low income level: 1**
**Older age score: 8**


*Ealing*

**Ealing**

**Cases: 85.78**
*(per 1000 people)*
**Deaths: 1.70**
*(Per 1000 people)*
**Deaths/cases: 2%**

**Crime level: 8**
**Low education level: 3**
**Low income level: 9**
**Older age score: 1**

# ROUND 3

*"Aline entered the round with a poor track record for crime, education and income and this caused a higher case rate. However, her youth helped her and she fought back to win and beat Paul, who had a much higher death rate"*

*Redhill*
**Reigate and Banstead**

**Cases: 59.38**
*(per 1000 people)*
**Deaths: 2.47**
*(Per 1000 people)*
**Deaths/cases: 4.2%**

**Crime level: 5**
**Low education level: 2**
**Low income level: 2**
**Older age score: 4**


*Reading*
**Reading**

**Cases: 64.12**
*(per 1000 people)*
**Deaths: 1.55**
*(Per 1000 people)*
**Deaths/cases: 2.4%**

**Crime level: 7**
**Low education level: 6**
**Low income level: 6**
**Older age score: 1**

# ROUND 4

*“Their stats might look even but population density helped Carlos beat Roberto on this one, with Roberto suffering from a higher death rate overall”*


**Spain**

**Cases: 64,196**
*(per 1,000,000 people)*
**Deaths: 1,164**
*(Per 1,000,000 people)*
**Deaths/cases: 1.8%**

**Lockdown Stringency: 7**
**Population Density: 6**
**Development Index: 9**

**Italy**

**Cases: 44,330**
*(per 1,000,000 people)*
**Deaths: 1,320**
*(Per 1,000,000 people)*
**Deaths/cases: 3%**

**Lockdown Stringency: 7**
**Population Density: 8**
**Development Index: 9**

# Thank you

I AM SMARTICUS