Guide	More Metrics COVID-19 Dataset
Executive Summary	**Free COVID-19 risk factor and infection data modelled to local geographic areas** Since April 2020, More Metrics has made available datasets that estimates COVID-19 risk factors and infection rates across the UK at a neighbourhood level. These datasets contain 20 different measures of risk at a range of local geographies. Where appropriate the risk measures have been updated each week to provide time-series estimates. Free to use data has been made available at Ward, Parliamentary Constituency, Local Authority and Clinical Commission Group level. This data is aggregated from our more detailed COVID-19 datasets available at Output Area (OA) and Lower Super Output Area (LSOA) that provides risk estimates for 230k and 43k neighbourhood locations respectively across all parts of the UK. During the course of this work, the open source COVID-19 data we use to build models have improved significantly, enabling us to refine and extend our approach. The current status of the supplied data is as follows: • We build disaggregated infection rate models each week for all parts of the UK using the confirmed case time series data at local authority or equivalent (England Wales and Northern Ireland) and by Health Board in Scotland. • These models have been calibrated to the PHE antibody testing results that provide cumulative infection rate estimates in the adult population by Region in England at specific time points. Using this data we create a conversion table between deaths and infection rates and confirmed cases that we then use to calibrate infection rates in England. We apply the conversion data for England to other Countries in the UK to estimate infection rates there on a comparable basis. • Using these improved models we have calculated weekly estimates from 5th April onwards for all geographies down to Output Area level. Currently this provides two different infection rate estimates for 11 time points up to the 21st June for c230k OA locations providing over 5 million infection rate estimates in total. The two different measures of infection risk are “As Is” the actual infection risk at each time point and “Time Adjusted” the expected infection risk assuming that all parts of the UK had become infected at exactly the same time point. The “As Is” values are generally higher than the “Time Adjusted” values in Regions like London that were infected first. In Regions infected later (e.g. Northern Ireland) the position is reversed with the “Time Adjusted” values generally higher than the “As Is” values. • From a projection of the year to date infection rate estimates we find the expected future change in infection rates to the end of the wave. These estimates are updated weekly for all Output Areas across the UK to help identify any areas that may be at risk of a jump in cases. • We have added new disaggregation models at Output Area level for confirmed cases per 100k population and mortality rates (as a proportion of all-cause mortality) to supplement our infection rate estimates. The mortality models use the detailed COVID-19 death counts published by ONS from time-to-time at MSOA level for England and Wales as source data. **At-risk areas in a second wave** We have also turned our attention to identifying those local neighbourhoods that are most at risk from a second wave. To do this we have estimated the day number each local neighbourhood achieved an infection rate threshold of 1 in 1000 and profiled how the risk characteristics of neighbourhoods vary with day number. This profiling uses a combination of our COVID-19 risk dimensions to create a COVID Risk Map. This positions every Output Area in the UK on a 100 x100 grid that defines their dominant COVID-19 risk characteristics. Each of these 10,000 cells on the COVID Risk Map averages the data for only about 23 Output Areas so this provides a very detailed view of COVID Risk. We demonstrate how this categorisation of risk can be used to chart the progress of COVID-19 across the UK by neighbourhood type from its outset to the current day.

	To complement this very detailed view of risk characteristics, we have developed a higher level segmentation that assesses the "second wave risk" for every Output Area using a 3 x 3 Risk Matrix. Each OA is assigned to one of the nine cells in the risk matrix based on the level of infection to date ("High", "Medium" or "Low") and the future risk of infection ("High", "Medium" or "Low") should a new outbreak occur. The calculation of these cell positions is found directly from the "As Is" and the "Time Adjusted" values for the year-to-date and future dimensions respectively. To support the wider analytical community investigating COVID-19, we are making our datasets at Ward, Parliamentary Constituency and CCG level freely available under the Creative Commons Attribution-Non Commercial 4.0 International Licence. For this study we have only used aggregated open-source data, which means that there are no GDPR implications clients need to be concerned with when using our COVID-19 datasets.	
Abbreviations	BMJ	British Medical Journal
	CCG	Clinical Commissioning Group
	OAC	Output Area Classification
	ONS	Office of National Statistics
	UTLA	Upper Tier Local Authority

Overview of the More Metrics COVID-19 Data	More Metrics provides innovative data solutions for a wide range of business and market sectors, from financial services to charities, energy providers and retailers. The COVID-19 dataset is a multi-dimensional dataset that assesses all neighbourhoods across the UK for risk using a set of rankings and modelled estimates. The data is available at a very detailed level (Output Area) with aggregations at higher level geographies. Our free to use data is made available at Ward, Clinical Commissioning Groups (CCG), Local Authorities and Parliamentary Constituencies and is supplied in a single Excel Workbook. The data we have used for this exercise is our existing More Metrics datasets derived exclusively from aggregated Open Source data. There are no GDPR implications that users should concern themselves with as we have not used any personal data or PII data in creating this output. The data series for risk measures are ranked versions of our modelled data by percentile, with 1 = lowest risk and 100 highest risk for Output Areas, LSOAs and Wards. The rank scale is from 1 to 20 in the case of Parliamentary Constituencies, Local Authorities and CCGs. We use these risk rank variables to visualise clusters of Output Areas on a "COVID Map" that have particular combinations of risk characteristics. We show that there is a clear relationship between the date COVID-19 started to emerge in different areas of the UK that is determined by particular neighbourhood risk characteristics. Infection rates are calculated using disaggregation modelling. All of our datasets include our estimated values for the "as is" infection rate each week from the 5th April 2020 onwards. A second "time adjusted" estimate of COVID-19 infection rates is also included. The time adjusted values are calculated on the assumption that all neighbourhoods in all parts of the UK are infected at the same time point. We use the combination of the "as is" and "time adjusted" infection rate estimates of infection rates to assign every Output Area in the UK to a 3 x 3 risk grid using a "High", "Medium", "Low" classification for the two measures. To complete the set of infection rate measures, we include an estimate of the future infections that will occur in phase 1 as a proportion of infections to date. This forward projection is calculated by curve fitting and extrapolation of the year to date data. For comparison purposes we also include an estimate of the R value for infections. This is estimated at the parent geography level only and is not disaggregated[1]. Impact outcomes are calculated for deaths and confirmed cases per 100k population and are disaggregated to provide estimates for all geographies. These outcome measures provide a better estimate of the actual impact COVID-19 has had on health care resources year to date and are better measures to use in this regard than infection rates on their own. Infections in a neighbourhood that is healthy and young will have a lower impact on the NHS compared to a neighbourhood with the same level of infection with residents who are older on average and who are in less good health. The outcome measures account for these age and health effects to give a truer estimate of the impact on the NHS. In total we provide 20 risk measures which we have grouped into six over-arching dimensions. These dimensions are as follows:

[1] It should be noted that the R-value and our forward projection of infections cannot be compared directly to each other. The R-value is not affected by the level of infection in a location whereas our forward projection reflects both the R-value and the absolute level of infection at any given point in time.

Risk ranking: Age and Household	This is aimed at identifying locations with a higher proportion of older people, those living in larger households, and those living in small spaces with a high number of residents per room. The five measures that sit within this dimension are • **All Age Risk** derived from ONS Age data weighted by COVID-19 death rates by age band. Includes communal residents. • **Household Age Risk** is derived from ONS Age data weighted by COVID-19 death rates by age band. Includes household residents only. • **Room Risk** is derived from ONS data for the number of household residents divided by the number of rooms in residential properties. This ratio gives an indication of overcrowding within properties. • **Resident Risk** is derived from ONS data for the number of household residents divided by the number of residential properties. This ratio gives an indication of household size with larger households having greater risk of catching COVID-19 from others in the household. • **Travel to Work Risk** is derived from ONS data for the means of travelling to work. Modes of transport that involve being in close proximity to other travellers (e.g. train and bus) are given a high-risk weighting with those with no contact (e.g. cycling) a low risk weighting. People who work from home or are economically inactive also receive a low risk weighting. The weighted average for all modes of transport is used to calculate the risk rankings of neighbourhoods.
Risk ranking: Mortality and Co-morbidity	This is aimed at identifying locations with a higher proportion of the population who have high health risk factors. All of the variables used for these rankings are age adjusted. The three measures that sit within this risk dimension are • **Mortality Risk** is derived from a More Metrics disaggregation of ONS published population death counts • **Obesity Risk** is derived from a More Metrics disaggregation of PHE population overweight proportion • **Smoker Risk** is derived from a More Metrics disaggregation of PHE smoker proportion and ONS lifestyle data.
Risk ranking: Economic Resilience	This is aimed at identifying those locations with low wealth and low Incomes before the COVID-19 outbreak who have fewer financial reserves to call on during the lockdown. In addition, we have analysed those neighbourhoods that are most likely to suffer additional changes because of potential financial hardships caused by the lockdown. This has a differential impact on those working in particular sectors of the economy defined by combinations of Industry Sector Risk and economic activity. Whilst these neighbourhoods may have good levels of wealth and income prior to the COVID-19 outbreak, they may be suffering a large drop in income during the lockdown and have to cut back or dip into savings to cover the gap. The three measures that sit within this risk dimension are: • **Income Risk** is derived from a More Metrics disaggregation of ONS Annual Survey of Hours and Earnings (ASHE) data • **Wealth Risk** is derived from a More Metrics disaggregation of Inland Revenue counts of estates subject to Inheritance Tax • **Employment Risk** is derived from a More Metrics imputation of Industry cross-tabbed with economic activity status (employed, self-employed, inactive) cross tabbed with hours worked (part time, full time). A subjective risk value of 1 (low) to 10 is attached to each combination of Industry x Economic Activity x Hours worked to reflect the impact of the lockdown on different groups. The weighted average of the risk value is calculated at Output Area level and converted to a percentile risk.

Risk ranking: Engagement	This is aimed at identifying those locations that may be less concerned and / or less well-informed about COVID-19 and its impacts. The risk is these neighbourhoods may pay less attention to the advice from Government resulting in higher infection rates. We have used out analysis of UK parliamentary petition data (pre the COVID-19 outbreak) to estimate these risks. For this analysis we have adjusted for local age profiles and Country to account for differences in participation rates caused by these two confounding factors. The two measures that sit within this risk dimension are: • **COVID-19 Engagement Risk** is derived from a More Metrics disaggregation of a basket of petitions relating to health, environment and education factors that demonstrate a high-degree of empathy with vulnerable groups. Those locations with low levels of engagement with these particular petitions are viewed as being of relatively higher risk. • **Overall Engagement Risk** is derived from a More Metrics disaggregation of all petitions. Those neighbourhoods that have a relatively low overall engagement in e-petitions may be less well informed on COVID-19 advice from Government and are viewed as being of relatively higher risk.
Risk ranking: COVID-19 Infection Rates	COVID-19 infection rates have been estimated by More Metrics on a best efforts basis. The method used to do this is outlined later in this document. The four measures that sit within this risk dimension are: • **COVID-19 infection rate "as is"**. This is an estimate of the cumulative infection rate defined as the proportion of the adult population that has been infected at some point with COVID-19 up to and including a particular date. The estimates are calculated weekly starting on the 5th April 2020 to show how the cumulative infection rates are changing over time.. • **COVID-19 infection rate timeline adjusted.** This is a cumulative infection rate estimate adjusted for different timelines and is calculated on the assumption that every part of the UK started to become infected at exactly the same time point. The overall level of infection is scaled to be roughly the same as the average "as is" value. This standardised measure is also updated weekly to track the progression of cumulative infection rates over time. • **Future Cases to Current Cases Ratio.** This ratio estimates the number of future COVID cases (infections) as a proportion of total cases (infections) to date. This is therefore a measure of the future COVID -19 risk for neighbourhood locations. A ratio value of 1 indicates there are estimated to be as many future cases to come in this location as there have been in total to date. A ratio of 0.1 indicates that future cases are estimated to be only 10% of cases seen so far. The source data used for this analysis is the published time series data for confirmed cases at higher geographies (e.g. LTLA, Health Board). We curve fit to this data to estimate the future trajectory of cases to calculate the future ratio value. These ratio values are then disaggregated and re-aggregated to obtain estimates at the geographies of interest. • **Average Daily Infection Rate or R-values.** These R values have been calculated at various time points at the "parent" geographic level where confirmed COVID-19 cases statistics are published. This is at LTLA level for England and Wales, Local Authority Area for Northern Ireland and at Health Board for Scotland. These values are not estimated at geographies below the published level, but where a lower level geography straddles more than one "parent" level, an average value is calculated weighted by population. We have used a calculation of RADIR values based on that described in the paper authored by Mike Stedman (Res Consortium, Andover) et al: https://www.medrxiv.org/content/10.1101/2020.04.20.20072264v1.full.pdf.

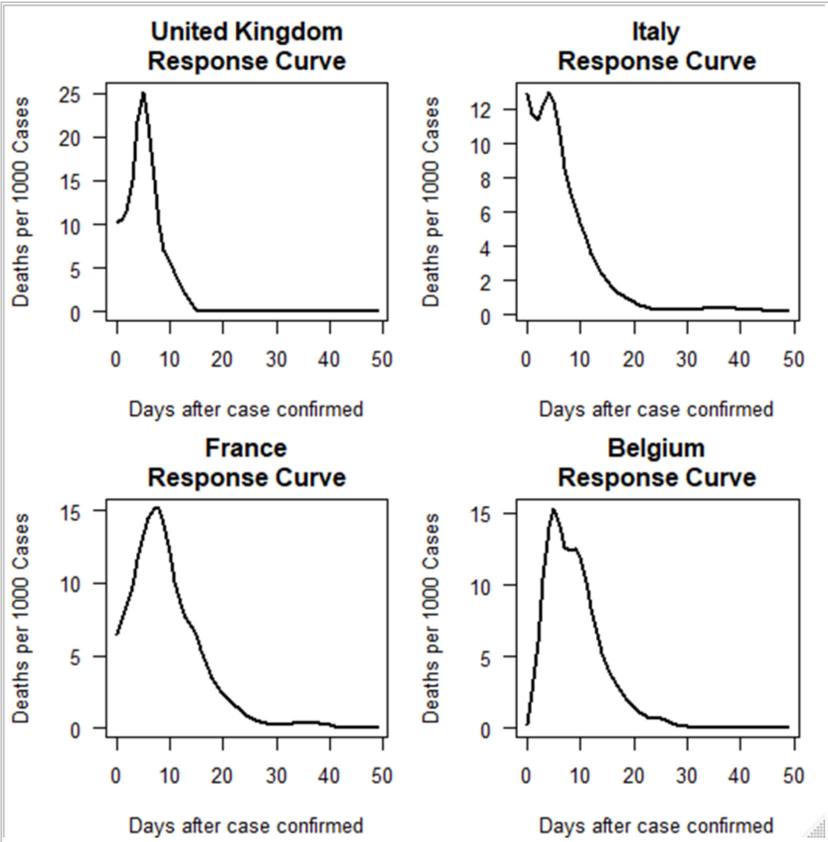
Risk ranking: Impact analysis for the NHS	The dataset includes some additional analysis that estimates the hospitalisation and mortality risk from COVID-19 at a neighbourhood level. The three measures that sit within this risk dimension are: • **COVID to All-Cause Death Ratio.** This estimates the number of COVID deaths as a proportion of deaths from all causes. This has been modelled from published ONS death data at middle super output area (MSOA) for England and Wales. The data is disaggregated to Output Area level. Estimates at other geographies are calculated from an aggregation of the Output Area values. The use of a ratio means that the figure takes account to some extent of the age profile and morbidity profile of the local population. The key driver of this ratio is therefore expected to be the underlying COVID-19 infection rate in the local population, providing an un-calibrated cross-check to our estimates derived from confirmed cases. Estimates outside of England and Wales have been estimated from the regression model coefficients. • **Confirmed Cases per 100k.** This estimates the confirmed cases per 100,000 population down to Output Area level from data published by the various national statistical bodies (e.g. ONS) and health authorities (e.g. PHE) at Local Authority level (or equivalent outside of England). Values at other geographies are calculated by aggregating the Output Area estimates. This measure provides a useful estimate of the stress placed on the NHS by COVID-19 during wave 1 as confirmed cases lead onto spells in hospital for a proportion of cases. However, it should be noted that because the average levels of testing are not the same across the constituent countries in the UK, care needs to be taken when comparing this measure between countries. Our infection rate estimates are calibrated across countries based on death rates and are therefore probably more reliable in that respect. • **Confirmed Cases Age Standardised.** This alternative estimated value for confirmed cases per 100k calculates the risk of someone being a confirmed case relative to the UK average for their age last birthday. This is a useful measure of the future burden on health services that could occur should an outbreak occur in a given location as it estimates the relative likelihood of someone being infected with COVID-19 displaying symptoms that result in a positive pillar 1 test result. The absolute number of cases that would occur given an outbreak will depend on additional factors, particularly the age distribution in the location. But for two locations with the same age distribution, this measure will identify the one at greatest risk of generating cases that need medical intervention.

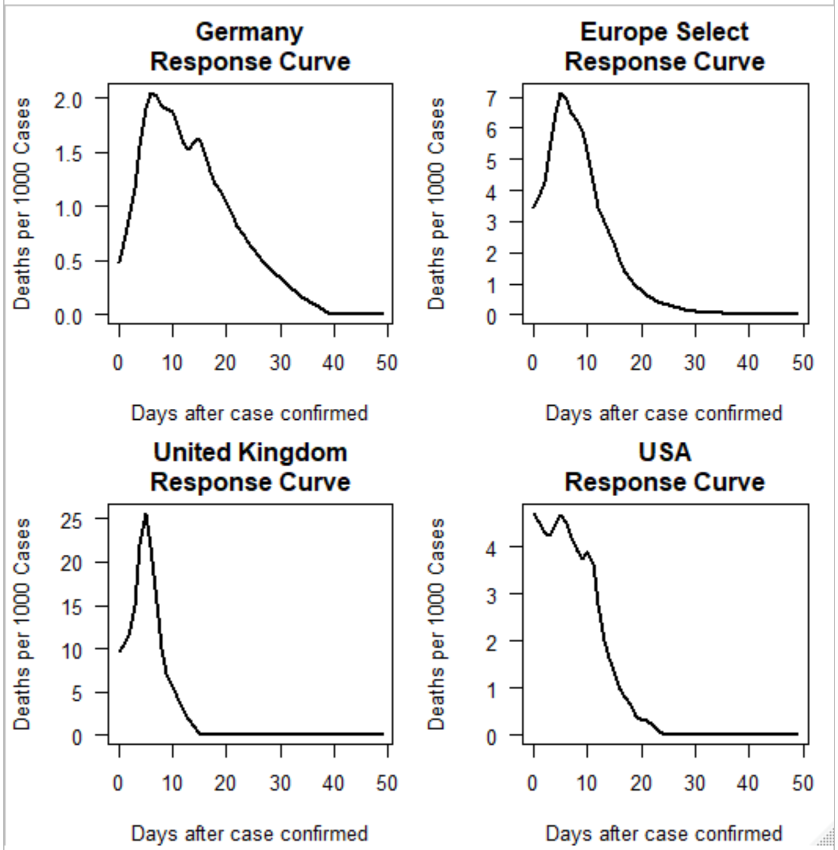
The method used to estimate UTLA COVID-19 infection rates	The analysis we have undertaken to investigate associations between COVID-19 daily infection rates and neighbourhood characteristics requires us to estimate COVID-19 infection rates at increasingly localised geographies, starting with published data at a national, Regional and local authority (or equivalent)[2], then estimating the infection rate at more localised levels using disaggregation (see below). The national and local authority infection rate estimates are obtained at weekly time points from the 5th April 2020 onwards. Infection rates are estimated from the cumulative number of confirmed cases and deaths using a method that takes account of different levels of testing in different Countries. We have used publicly available data about COVID-19 death and infection rates (from antibody testing) for England in order to calculate the necessary conversion rates from confirmed cases to infections for England, taking account of time lags between cases and deaths and the age profile of infections. Having obtained a set of conversion rates for England we apply these to other countries in the UK to calculate infection rates on a consistent basis. There are a number of key elements to the calculations needed to undertake these conversions accurately which split into a sequence of sequential steps which are now described.
	1. Account for the lag in deaths compared to cases (a global overview). It is well understood that there is a lag between the recording of deaths and COVID-19 confirmed cases with one UK based study indicating that most COVID-19 patients recover or die within a period of up to 17 days after admission to hospital. It is therefore logical to assume that deaths for cases primarily occur after testing positive for COVID-19 when a subset of those tested, who are deemed to be in need of medical attention, are admitted to hospital for observation and treatment. If a mass testing programme is underway (e.g. Germany) then it is likely that many more cases are detected early and the response curve will be elongated. This situation contrasts with countries which are slow to set up testing programmes (e.g. Italy, England), and are then overwhelmed by hospital admissions. In these locations test results may only become available close to or after death has occurred, shortening the response curve timeline. Our analysis of time series data for cases and deaths published by Our World in Data (OWID)[3] quantifies the lag between deaths and cases for a wide range of countries, focussing on the first 50 days of the pandemic in each location where reliable response curves can be calculated from this aggregated data. The R code we have developed for this analysis is available on request and the results using this code are shown below starting with selected European countries[4], including those that were some of the hardest hit. Whilst the scale and shape of the response curve linking deaths to cases varies considerably, the evidence of a lag is seen consistently. In the case of these four countries, the average lag extends over a period of between 15 and 25 days with a prominent peak between 7 to 10 days for all but Italy. We see a shortening of the response curve for the UK which corresponds to the country with the highest peak in the response curve death rate as a result of very low testing rates at the outset of the pandemic.

[2] The published data we use for Scotland is at Health Board level

[3] We have used "Our World in Data" resources for this analysis. The source data and license details can be found here: https://github.com/owid/covid-19-data/tree/master/public/data

[4] "Europe Select" is an unweighted average of the response curves for the following countries: Austria, Belgium, Bulgaria, Czech Republic, Denmark, France, Germany, Greece, Hungary, Ireland, Italy, Netherlands, Poland, Portugal, Romania, Slovakia, Sweden and the United Kingdom. Spain is not included because the available time series data from OWID contains significant corrections.



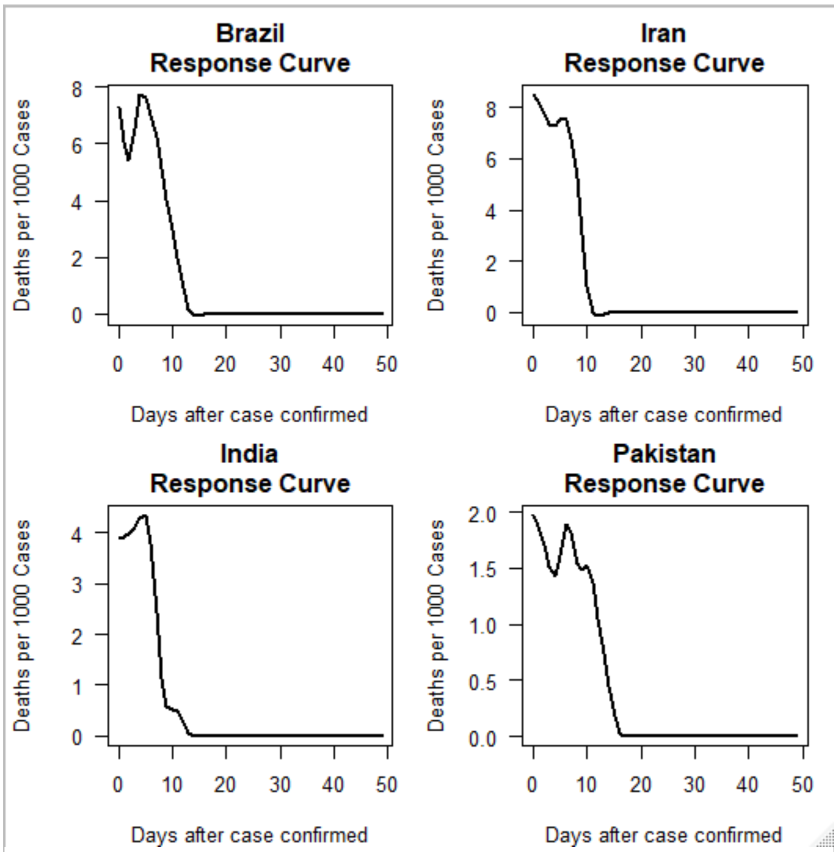


Italy shows a peak close to day zero which we attribute to the initial explosion of cases in the Northern Region which overwhelmed local medical care resources resulting in the triage of testing to the most critical patients where deaths occur just before or shortly after COVID-19 infection is confirmed.

If we look at the response analysis for Germany we can see the beneficial impact of mass testing from the outset of the pandemic. This shows an elongated response curve extending out to 40 days and a low death rate peak. This contrasts with the rest of Europe more generally and the UK in particular with its much shortened response curve and higher death rate peak.

The USA shows a different pattern to Europe as a whole that is more like Italy with a peak that is close to day zero. Like Italy the initial infection exploded in a specific area of the country (New York and new Jersey States) which may be one reason for this response curve shape.

However in the case of the USA, we consider that this early peak could also be caused by a “two tier” health system. This hypothesis assumes that the COVID-19 at-risk-patient-groups in the USA are more likely to have limited access to primary health care, meaning they will be tested later on average than those in Europe. This is likely to be when they are seriously ill and require hospitalisation, shortening the overall time seen between testing and death for this group of patients.



In this respect the USA response curve shows that it has similarities to other nations with less well developed health-care systems. Here we show plots for Brazil, Iran, India and Pakistan to support this observation. The key difference between these four countries and the USA is that the overall response curve is shortened to between 10 and 15 days, whereas the USA tail extends out to day 20, which we attribute to a higher proportion of confirmed cases in the USA that are well-insured and receive world class medical treatment resulting in earlier detection and extended hospital stays for this sub group.

2. Estimating accurate death rates from confirmed cases for the constituent countries of the UK

We can obtain an accurate fit to the number of deaths recorded on a cumulative basis for each nation in the UK by applying an appropriate response curve to the number of reported confirmed cases each day. To achieve the best fit we vary the weighting applied to the response curve over time to account for increases in testing volumes during the course of the pandemic. This analysis provides us with a conversion rate between confirmed cases and deaths at every time point that deals with the time lags involved. The resulting confirmed case to death conversion rate varies nation by nation because of the different approaches and levels of testing adopted under the devolved powers for each country.

3. Calculating infection rates using an assumed infection fatality rate (IFR)

If we assume an average infection fatality rate for COVID-19 that is the same across each country, it is straightforward to convert deaths (adjusted for lags) into infections at any time point. For example if the average infection fatality rate from COVID is assumed to be 0.5%, each death represents 200 infections. This then allows us to calibrate infection rates between countries within the UK assuming that the total number of infections is a multiple of the total number of deaths (adjusted for lags) rather than the total number of confirmed cases.

Having established an overall estimate for the infection rate at a country level we can then use the cumulative confirmed case counts for local authorities within each country to estimate how the infection rate has changed over time at a sub-national level. The assumption is made that testing protocols and levels within countries is broadly consistent unlike the situation between countries.

When we started our analysis we applied this simplistic approach using the widely accepted figure for average COVID-19 death rates of 0.66%. We did this at the time because no reliable estimates of actual infection rates were then available.

4. Calculating a better estimate for the infection fatality rate (IFR)

Public Health England now publishes infection rate estimates obtained from an analysis of antibody tests in the general population (sampled from healthy adult blood donors). The results are detailed in the PHE weekly surveillance reports in the section entitled: "Sero-prevalence epidemiology, England".[5] This is an extremely valuable source of insight into infection rates that provides data at a Region level within England as well as important demographic detail. For example in the week 24 report PHE note the following about the age profile of infections:

> *"Age specific prevalence estimates have changed over time with prevalence notably higher in the young adults in those areas that experienced the highest incidence in the earlier weeks of the outbreak. Over time however the prevalence in older adults increased more suggesting that this age group were being affected later. These patterns may reflect differences in behaviour and mixing patterns in the different age groups."*

Armed with this PHE analysis on infection rates in the general population we have undertaken a more rigorous analysis of the link between observed death rates and infections for English data. The aim of our analysis is to use the findings from different sources to obtain a set of conversion tables we can apply generally to the published data on cases and deaths to obtain reliable infection rate estimates at sub-national level. An important part of this analysis is to ensure that the resulting conversion tables are consistent with a wide variety of different sources as far as this is possible, thereby increasing confidence that our infection rates are as robust as they can be.

The key sources we use to estimate and / or assess our conversion rates include the following:

1. The observed confirmed case and death data in England.
2. PHE infection rates from antibody testing for England.
3. The Imperial College assumption that the Infection Fatality Rate (IFR) from COVID-19 if left to run unchecked (do nothing scenario) would be 0.9% resulting in circa 500k deaths in the UK[6]
4. The many reported statistics for COVID-19 cases and deaths by age band that shows the extremely large variation in Case Fatality Rates (CFR) by age, with younger people facing very small risks from COVID-19 compared to people over retirement age. See the work published by Professor David Spiegelhalter for an analysis of this important issue.[7]
5. The University College London (UCL) COVID-19 Social study findings for Compliance with Guidelines by age group that shows complete compliance increases with Age. The youngest age group 18-to 29 is less likely to comply (at c50% at the time the analysis was done but now at c40%) compared to the oldest age group 60 and over (at c70% when the analysis was done but now at c60%).[8]
6. COVID Symptom Study developed by health science company ZOE and analysed by King's College London, that shows a fall-off in Symptomatic COVID for ages 60 and above[9]

---

[5] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/891721/Weekly_COVID19_Surveillance_Report_-_week_24.pdf

[6] http://www.imperial.ac.uk/news/196234/covid-19-imperial-researchers-model-likely-impact/

[7] See https://medium.com/wintoncentre/what-have-been-the-fatal-risks-of-covid-particularly-to-children-and-younger-adults-a5cbf7060c49

[8] See https://www.covidsocialstudy.org/results

[9] See https://covid19.joinzoe.com/

We will now outline the process we have used to calculate our conversion tables for estimating infection rates from death counts. The first stage in the process is to create a segmentation of the general population that helps to explain important age differences. In particular, the UCL social study finding that shows compliance increasing with age (suggesting infection rates should fall with age all other things being equal). To map our risk rank data to the UCL research findings we decided it would be useful to split the UK population into three broad categories as follows:

- ***Minglers.*** People with the highest risk of infection who through inclination or circumstance have significant contact with others. This group includes those who are socially and economically active and who use public transport. Health workers and key workers would be included in this category as well as those living in large family groups that contain people less inclined to follow the guidelines. We expect this category to be over-indexed on younger ages and also on larger households that contain resident workers in key roles.
- ***Non-Minglers.*** People with the lowest risk of infection who are able to self-isolate and are motivated to follow the guidelines (e.g. the worried-well who are not economically active). This group includes those who live independently and are able to avoid unnecessary contact. Healthy empty nesters and furloughed couples with school aged children will also fall into this category. We expect this category to be over-indexed on middle and older ages, particularly for those that have larger houses and gardens and who have the financial means to minimise their risk profile without suffering undue hardships.
- ***Enforced-Minglers*** People with an intermediate risk of infection who would ideally self-isolate, but who are not able to do so because of their circumstances. This group is primarily made up of people with care needs resulting in the oldest age groups being over-indexed for this category.

We have used the 2011 census data for care-home residents by age as a proxy for the Enforced-Minglers. We have arbitrarily doubled these proportions to take account of the number of people receiving care at home, which we therefore assume is equal to the number in residential care and with the same age profile which may need further refinement.

We then split the remainder of the population between Minglers and Non-Minglers. To do this we rank order the mingler proportions by age using the combined average of our risk indices for Travel, Room and Resident.

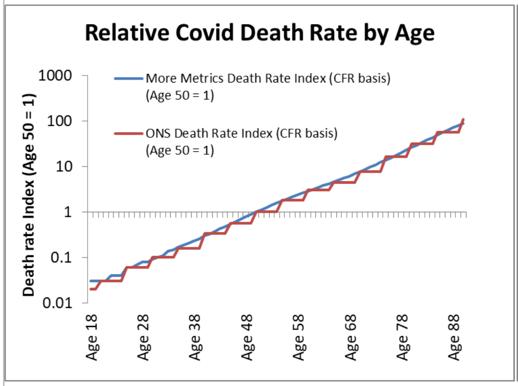
The chart below left shows how this combined risk rank measure varies by age when weighted by the proportion of the Output Area population in each single year age category. It shows the expected fall with increasing age

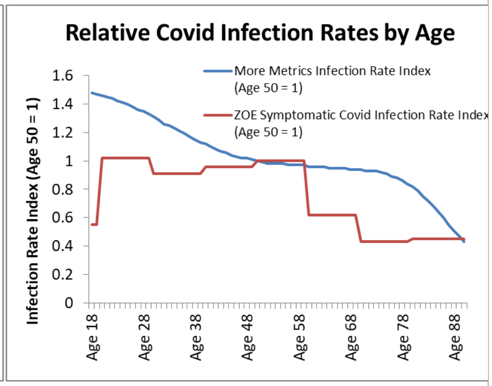
This is supporting evidence that it is harder for young adults (on average) to fully comply with the guidelines because of their circumstances. They generally live in more densely populated households and neighbourhoods, and have a greater reliance on public transport.

We convert this rank measure for the Mingler proportion to absolute values by calibrating to the published data for UCL proportions by coarse age band for those who do not fully comply with guidelines. The remaining fraction is an estimate of the proportion of the population who fully comply with the guidelines (Non-Minglers). The resulting population proportions by single age are shown in the chart below right.

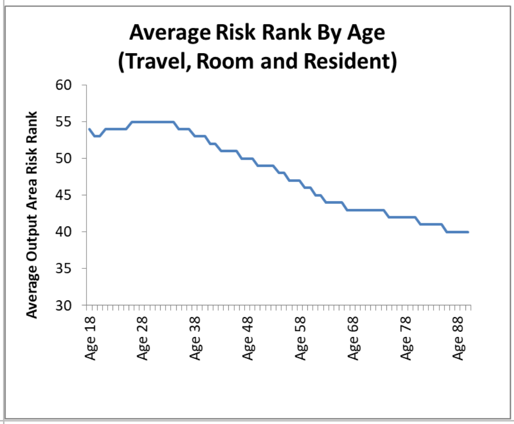
Using these proportions we run an iterative process to calculate Infection Rates for each risk category that simultaneously achieves the Imperial College Infection Fatality rate (IFR) of 0.9% under the do nothing scenario (i.e. assuming everyone is infected) and also meets the observed Case Fatality Rates (CFR) reported by ONS for England and Wales. The CFR values differ from the IFR values because testing does not identify 100% of cases and lockdown rules have been designed to protect the elderly to ensure that infection rates are also not uniform across age bands.

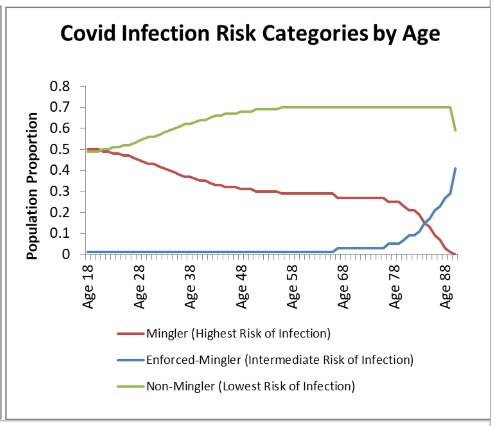
We therefore allow the average Infection rate for Minglers and Non-Minglers to vary as necessary to achieve convergence. The infection rate for Enforced-Minglers is arbitrarily set at three times the Infection rate of Non-Minglers when we do these calculations. The output from this process is shown in Appendix 1 and in the graphs below.





The ONS measured variation in death rates by age (Case Fatality Rate, CFR basis) is closely matched by our results (see chart above left). This is achieved by varying the infection rate by age, with younger residents estimated to have higher levels of infection (see chart above right). Our infection profile by age does not follow the results from the ZOE app very closely, but are directionally similar. We note however that the ZOE app measures people who identify as having COVID-19 symptoms and therefore the difference might be explained if younger people are more likely to be asymptomatic.





5. Obtaining sub-national infection rates using the conversion table data

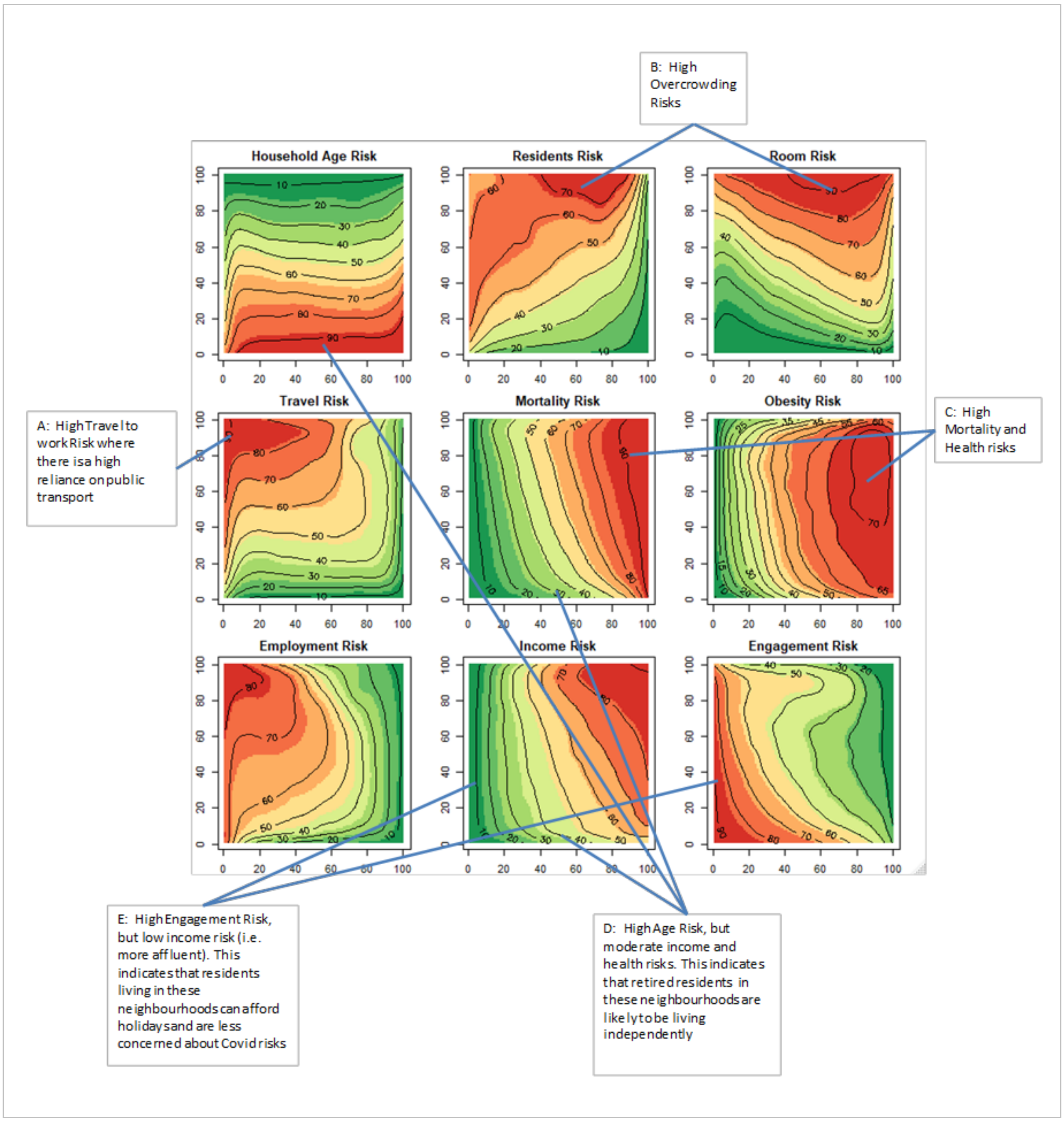
The conversion table data for England is used to estimate infection rates at a sub-national level for 366 locations across the UK. Sub-national is local authority (or equivalent) in England, Wales and NI and Health Board in Scotland.

For each sub-national level we first convert the observed number of confirmed cases into an estimate of the number of deaths accounting for time lags using the appropriate response curve and scaling factor. The scaling factor is set nationally to ensure that the sum of deaths at a sub-national level found from converting confirmed cases to deaths agrees to the published value for deaths at a National level.

The number of deaths at sub-national level is then converted to the number of infections. The conversion rate for deaths to infections is calculated for each location from the age distribution in each location and our own estimates for how the infection rate and death rate index varies with age overall. The effect of this calculation means that locations that are younger on average will be estimated to have a higher infection rate to death ratio, for a given case rate. This is because, as previously shown, we estimate that younger people are more likely to be infected and also have a much lower death rate compared to older people.

Local Infection Rate estimation using Disaggregation	To obtain localised estimates of infection rates we apply our disaggregation method to obtain modelled estimates of infection rates at Output Area level. Once the Output Area estimates are obtained we re-aggregate these local estimates back to other geographies of interest, namely Ward, Parliamentary Constituency, Local Authority and CCG. Disaggregation involves building a regression model iteratively to apportion the sub-national infection rates across neighbourhoods based on their characteristics. We have found that our disaggregation method gives reliable local estimates when the data available is of high quality and well distributed at higher geographies. In this case, the modelling dataset used for disaggregation covers all parts of the UK and contains data for 366 individual sub-national locations which enables us to build fit for purpose disaggregation models. As previously mentioned, we calculate two sets of disaggregation models which are the "As Is" and "Time Adjusted" variants. The reason for doing this is that different parts of the UK started out at different time points on the infection curve relative to the imposition of lockdown rules, with London days ahead of other areas of the country. We therefore want to be able to identify changes in infection rates that are primarily due to timing differences separately from the differences in infection rates attributable to socio-demographic factors. To account for the timing issue we use daily case growth rate estimates for each sub-national location to calculate how long ago each was at a very low infection rate of 0.1% (1 in a 1000). This allows us to place all of the sub-national locations along a time line. We then use local smoothing to estimate how the day number of locations below sub-national level varies. This is achieved by smoothing the estimates centred on Lower Super Output Areas, for a rolling average population of 50k. This process smooths the day number estimate using data for all nearby locations. The smoothing process is done repeatedly to obtain a stable set of values for every LSOA that correctly aggregates back up to the sub-national values calculated at the outset. Time variables and Country categorical variables are included in the disaggregation models to account for the timing effect separately from the variations related to local neighbourhood characteristics. We model the "As Is" values with all terms included. We then calculate a second "time adjusted" infection rate with the time and Country variables set to the UK average. The time-adjusted estimate is therefore the best one we have for investigating the impact of local characteristics on infection rates on a uniform basis across the UK.

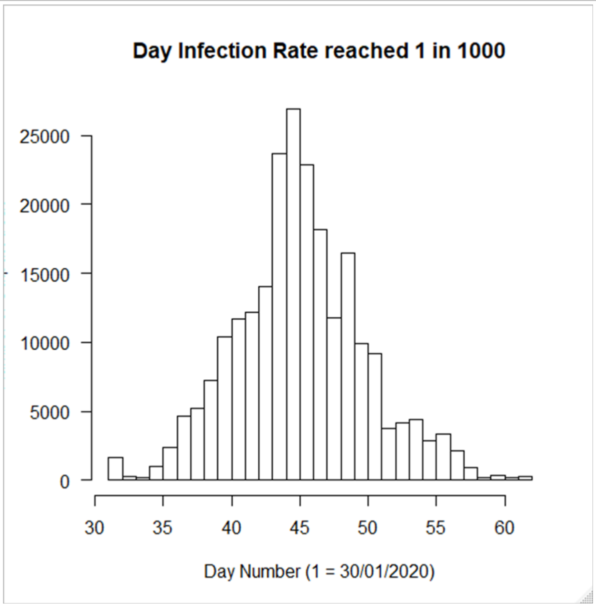
Neighbourhood categorisation for COVID-19 Risk Assessments	There are many existing ways of categorising neighbourhoods that are being used to help throw light on COVID impacts. These include identifying neighbourhoods that have higher levels of deprivation and / or a higher proportion of BAME ethnicities which are observed to be more adversely impacted in terms of infection rates and adverse outcomes. More generally the use of commercial socio-demographic classifiers (e.g. ACORN, CAMEO) and open source datasets such as the ONS Output Area Classification (OAC) categories are likely to be predictive. Commercial offerings that are more specific (e.g. those linked to health outcomes) may add further value. Against this backdrop, we will now show how our datasets provide something uniquely different. Our data is designed to complement all of the other available sources when assessing Wave 2 COVID-19 risks at a local level. The results we will review in this section cover: • **COVID Risk Mapping:** Categorising local areas by their combination of More Metrics COVID Risk Ranks. • **Day Number Analysis:** Categorising local areas by when infections started (early to late) and then using COVID Risk Maps to show that timing in relation to lockdown is a major factor in how different neighbourhoods are affected. • **Wave 2 COVID Risk Matrix:** Categorising local areas as they approach the end of Wave 1 by the level of infections they have achieved year to date and their potential for future high rates of infection going forward to create a summary 3 x 3 Risk Matrix.
	1. COVID-19 Risk Mapping The COVID-19 Risk map provides a visualisation of COVID-19 risks across different dimensions. The map is created using a selection of our risk ranks to undertake a 2-dimensional factor analysis. The factor coefficients are then used to position individual Output Areas on a 100 x 100 grid with an average of 23 Output Areas at each point on the grid. <u>The grid is oriented so that age-standardised health generally worsens from left (more healthy) to right (less healthy) and neighbourhood average age decreases from bottom (old) to top (young).</u> The average value for each risk index is then calculated and risk contours are plotted. Visual inspection allows us to identify different areas of the risk map associated with patterns of risk across multiple dimensions. As we see boverleaf, this gives us a useful framework for analysing COVID-19 infection rates. High risk is red and low risk is green. In the example below we have identified five locations on the risk map that have specific risk characteristics as follows: • Zone A (Top Left Corner) is where Output Areas that have a high travel to work risk are situated • Zone B (Top, Middle / Right) is where Output Areas that have high overcrowding risks are situated • Zone C (Right Side, Upper/ Middle) is where Output Areas that have high mortality and morbidity risks are situated • Zone D (Bottom Middle) is where Output Areas that have high age risk, but moderate health and income risks are situated • Zone E (Left Side. Middle / Bottom) is where Output Areas that have high engagement risk, but la ow income risk are situated. Our Output Area dataset provides the grid location for every OA on the COVID Map which therefore gives a specific assessment of the balance of COVID-19 risks for each neighbourhood matched to postcode. This is ideal information for use when comparing risks between locations and for scenario planning, especially where other data sources are being used that are sparsely populated with data limiting the ability to generalise results.

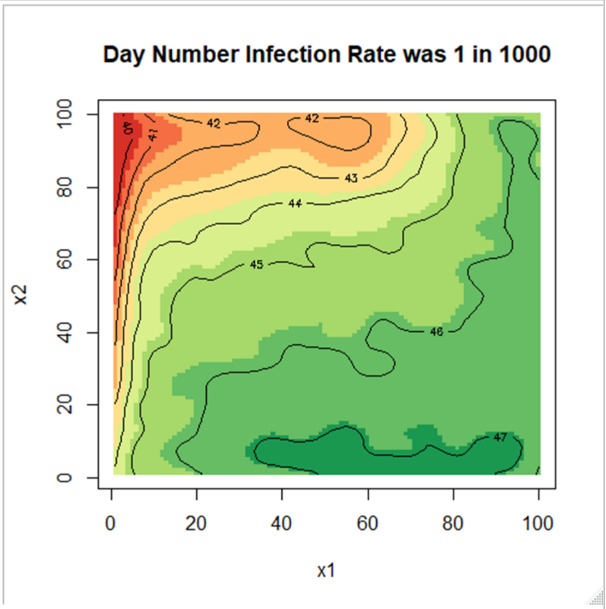

B: High Overcrowding Risks
Household Age Risk
Residents Risk
Room Risk
A: High Travel to work Risk where there is a high reliance on public transport
Travel Risk
Mortality Risk
Obesity Risk
C: High Mortality and Health risks
Employment Risk
Income Risk
Engagement Risk
E: High Engagement Risk, but low income risk (i.e. more affluent). This indicates that residents living in these neighbourhoods can afford holidays and are less concerned about Covid risks
D: High Age Risk, but moderate income and health risks. This indicates that retired residents in these neighbourhoods are likely to be living independently

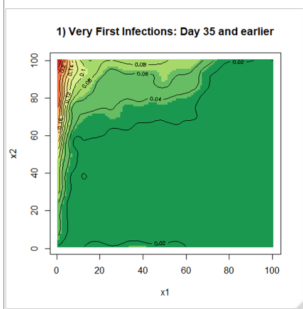
2. Day Number Analysis

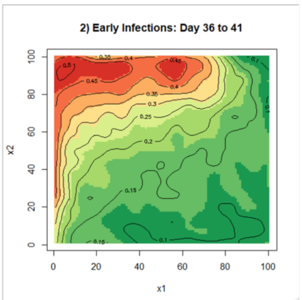
The day infection rates reached 1 in 1000 is estimated in our data for every LSOA across the UK. With one or two exceptions, infections started earliest in London and spread across England, with other parts of the UK following on. The chart on the left shows the day number distribution by Output Areas for the whole of the UK.

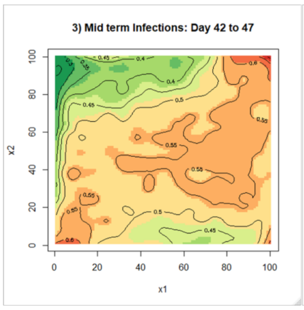
The early start in London meant that the infection took hold in some areas before Lockdown measures took full effect. Day number analysis using the COVID Map allows us to chart this progression over time. The overview chart on the right shows how the infection first took hold in younger neighbourhoods with high travel risk (top left corner). The bottom four charts show how the infection then progressed to the right and then down towards the bottom of the map where the age risk is highest, meaning that older neighbourhoods were infected last, presumably as a result of social distancing delaying its impact on the most vulnerable. This pattern of COVID-19 emergence in our data is entirely consistent with the observations by PHE about how infection rates measured by antibody testing have changed over time by age band.

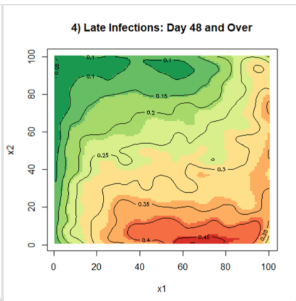












This analysis shows the critical importance of high travel risk neighbourhoods in "seeding" the infection into neighbouring areas. Areas with high overcrowding risks were then next to see a high growth in infections that then moved from there into most other areas. Neighbourhoods with older residents who were able to self-isolate early on were last to be infected. The decision to require face masks on public transport is fully supported by our analysis. Stopping infections spreading through strangers mingling at scale will be of critical importance in managing wave 2.

We can crosstab our day number analysis with the ONS Output Area Classification (OAC) categories to provide additional insight at a more local level. By way of example, the analysis below looks at London and Scotland, and charts where the infection started first by OAC neighbourhood type in these locations at the very outset. Scotland is about a week behind London, but there are similarities between the two locations in a number of characteristics that include residents born in the EU, students and multi-ethnic / multicultural neighbourhoods and Professional Service workers (perhaps reflecting locations over indexed on doctors and nurses?).

Using our most localised datasets, this analysis can easily be replicated for other parts of the UK areas and for crosstabs against alternative neighbourhood categorisations (e.g. by ethnicity and deprivation deciles). Also analysis teams with access to individual level data on infections can use our neighbourhood level datasets to help generalise their findings and to deal with data gaps.

London Top 10 (First to be infected)		
OAC Sub Category Name	Day Number	Number of Output Areas
2d2 - Highly-Qualified Quaternary Workers	36	1211
3b3 - Multi-Ethnic Professional Service Workers	37	874
3b1 - Striving Service Workers	37	1356
3d1 - New EU Tech Workers	37	1246
2d3 - EU White-Collar Workers	37	1171
3b2 - Bangladeshi Mixed Employment	37	652
2b2 - Multicultural Student Neighbourhoods	37	423
2a1 - Student Communal Living	38	28
3d3 - Old EU Tech Workers	38	1324
3d2 - Established Tech Workers	38	957
2d1 - Urban Cultural Mix	39	640

Scotland Top 10 (First to be infected)		
OAC Sub Category Name	Day Number	Number of Output Areas
2d3 - EU White-Collar Workers	44	41
4a3 - Commuters with Young Families	44	18
2d1 - Urban Cultural Mix	45	18
2b1 - Students and Commuters	45	1347
3c1 - Constrained Neighbourhoods	45	56
2c3 - Professional Service Cosmopolitans	45	1158
2b2 - Multicultural Student Neighbourhoods	45	81
4a1 - Social Renting Young Families	45	20
3c2 - Constrained Commuters	45	527
4c1 - Achieving Minorities	45	32
4c3 - Inner City Ethnic Mix	45	19

Note: OACs with less than 10 Output Areas not included in this analysis

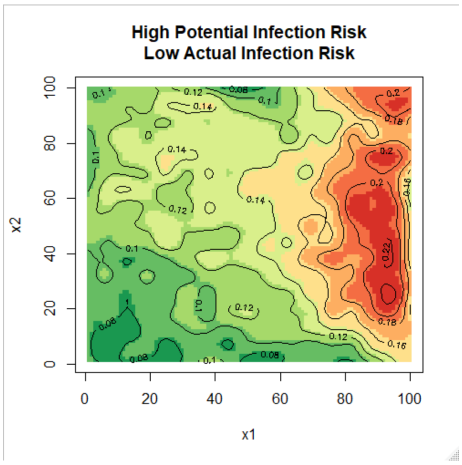
3. Wave 2 COVID Risk Matrix

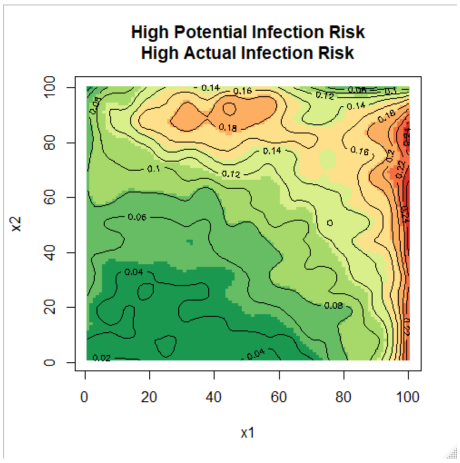
The More Metrics Risk Matrix places every Output Area in the UK into 1 of 9 cells on a 3 x 3 Risk Matrix that categorises the "Wave 2" risk for each neighbourhood based on the latest "As Is" and "Time Adjusted" infection rate estimates for each Output Area.

Using these measure the Actual and Potential Risk rank is derived to create the matrix detailed below

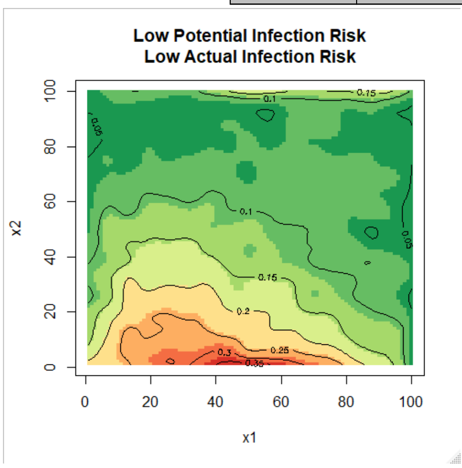
Potential Risk Rank ("Time Adjusted")				
	High	***Deprived neighbourhoods mainly outside England*** Highest 2nd wave risk with no HIT achieved anywhere?		***First infected, mainly in London*** Herd Immunity Threshold (HIT) reached for some, lowering 2nd wave risk?
	Medium			
	Low	***Rural neighbourhoods and Independent Retired*** Easier for residents to shield lowering risk?		***More Affluent Areas mainly in England*** Early Infection spike from returning skiers?
More Metrics Risk Matrix		**Low**	**Medium**	**High**
		Actual Risk Rank ("As Is")		

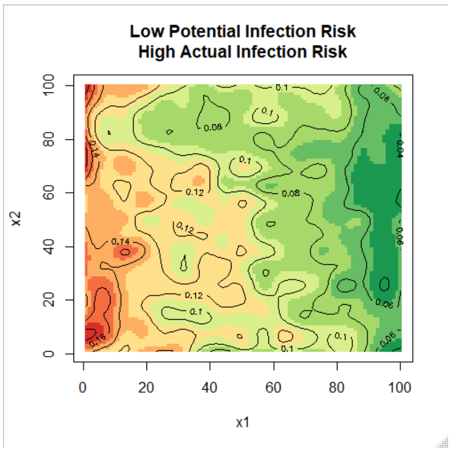
	The COVID Maps for the four corners of the Risk Matrix show that different neighbourhood characteristics are associated with different risk categories. The highest second wave risk is considered to be in the High potential infection risk / Low actual Infection risk cell of the matrix (top-left square). This is because these neighbourhoods have health characteristics associated with poor COVID outcomes combined with high infection risk should an outbreak occur. The low "As Is" infection rates in this Matrix cell also mean these neighbourhoods have little chance of having achieved Herd Immunity Threshold (HIT) for any sub population.





Potential Risk Rank ("Time Adjusted")				
	High	***Deprived neighbourhoods mainly outside England*** Highest 2nd wave risk with no HIT achieved anywhere?		***First infected, mainly in London*** Herd Immunity Threshold (HIT) reached for some, lowering 2nd wave risk?
	Medium			
	Low	***Rural neighbourhoods and Independent Retired*** Easier for residents to shield lowering risk?		***More Affluent Areas mainly in England*** Early Infection spike from returning skiers?
More Metrics Risk Matrix		Low	Medium	High
		Actual Risk Rank ("As Is")		





Using Our Data in the Real World	We will now consider three hypothetical case studies that show how our data might be used by different stakeholders to best manage the journey out of lock down. These examples are for illustration purposes only and are not comprehensive in scope.

Case Study 1

Local Council / Member of Parliament faced with a local lockdown

Burning Issue: How does a Local Council and the Members of Parliament that represent its residents work together to avoid the need for a new lockdown by taking the right pre-emptive actions?

Some suggestions: Good risk management and contingency planning requires access to reliable information focussed on identifying risks early that are shared across all relevant parties, which in the case of COVID-19 means everyone!

A useful analogy is how communities avoid the worst effects of bush fires. A multi-layered approach is needed. Prevention involves prioritising actions (e.g. clearing or damping down tinder) in the locations of greatest known risk to maximise effectiveness. Targeted communication is focussed on those neighbourhoods most at risk and on those individuals whose behaviour is most likely to spark an outbreak. Then there is an intelligent use of up-to-date satellite and weather data to enhance early warning systems so that vigilance activity is dynamically allocated to the highest risk. Finally should an outbreak occur, a multi-agency response swings into action quickly and decisively by executing a pre-planned response.

Why use More Metrics Data to support this activity? Our data is updated weekly and includes a forward projection of the COVID-19 infection risk that is simple to understand and interpret. Using this measure as at 21st June 2020 for Parliamentary Constituencies, we can identify those most at risk of continued infections that need to be particularly vigilant. This analysis is the equivalent of identifying those locations of greatest known risk (using our bush fires analogy) combined with updates that enhance early warning systems when used alongside other routinely collected data. This therefore directly supports a number of the critical requirements of a well thought through risk prevention strategy.

So how does this work in practice? In the table below we have used our free to use data to create a risk table for 650 parliamentary constituencies. The data in this table is sorted (highest risk first) by column E which estimates the percentage of the population yet to be infected based on current trends.

Parliamentary future risk		A	B	C	D	E
Parliamentary Constituency	**Pcon Code**	**AsIs Infection Rate 21st June 2020**	**Time Adjusted Infection Rate 21st June 2020**	**FutureCases To Current Cases Ratio 21st June 2020**	**R Value 14th June 2020**	**Future Infections (A x C)**
Ynys Môn	W07000041	7%	12%	40%	1.98	2.9%
Arfon	W07000057	10%	14%	26%	1.505	2.6%
Castle Point	E14000622	10%	9%	15%	0.958	1.6%
Rayleigh and Wickford	E14000888	10%	9%	16%	1.022	1.6%
Barnsley Central	E14000541	20%	18%	8%	1.131	1.5%
Leicester South	E14000783	16%	14%	9%	0.927	1.4%
Southend West	E14000957	11%	10%	13%	0.769	1.4%
Bedford	E14000552	21%	19%	7%	0.547	1.4%
Northampton North	E14000861	14%	12%	9%	0.902	1.3%
Leicester East	E14000782	18%	15%	7%	0.927	1.3%
Stalybridge and Hyde	E14000967	18%	16%	7%	0.63	1.2%
Manchester Gorton	E14000808	20%	18%	6%	0.716	1.2%
Ashford	E14000536	26%	23%	5%	0.779	1.2%
Wrexham	W07000044	10%	13%	11%	0.805	1.1%
Dover	E14000670	15%	14%	7%	0.922	1.1%
Leicester West	E14000784	17%	15%	7%	0.927	1.1%

We see that three constituencies in Wales have high future infection risk Anglesey (2.9%), Arfon(2.6%) and Wrexham (1.1%). Other danger signs for these parliamentary constituencies are:

- The high R Values for the local Authority Parent close to or above 1
- The high “Time Adjusted” infection rate values which lie above the “As Is” infection rate values indicative of areas still working their way through phase 1 before “steady state” is achieved under the historic lockdown rules. Not reaching “steady state” before lockdown rules are relaxed is particularly risky (as can be observed by the situation playing out in the USA).

The rest of the high-risk constituencies are situated in England. Three of these are based in Leicester (highlighted in blue) which, at the time of writing, is being considered an area requiring a local lockdown. This unfortunate situation confirms how critical it is to monitor the future risk continually using all available resources to pre-empt this kind of response when things go wrong.

In this context, our table shows a further 10 constituencies in England that are of similar level of risk right now to the Leicester constituencies. This should immediately prompt MPs and Councillors in these locations to be on high alert, calling on national and local resources to re-double their collective efforts to stop a “flare up”. Preventative action taken now in these “tinderbox” locations will have huge benefits in de-risking the course of the infection without the need for possible draconian action later with the unwanted knock-on effects this will have on the local economy and the morale of residents.

Targeted communication is an important component in this pre-emptive strike.

The messages therefore need to vary a lot, for example:

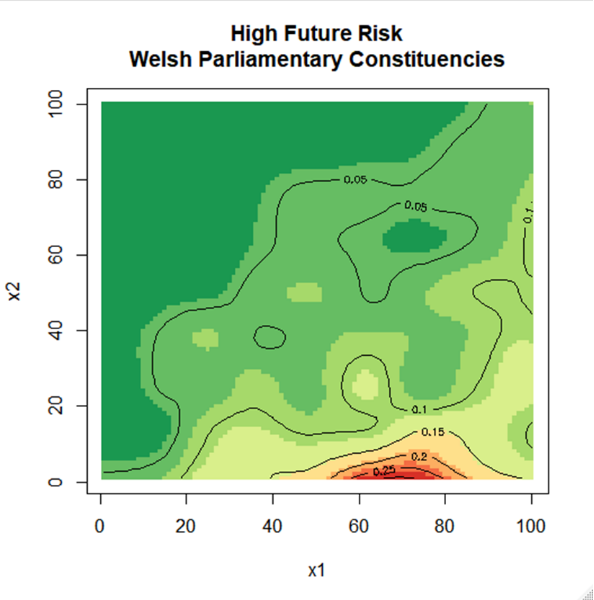
- By Location
- By Age Band
- By Business type
- By channel and message giver

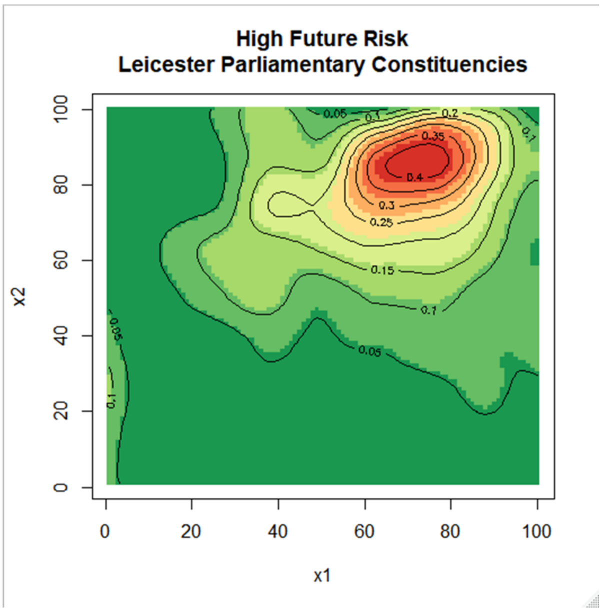
To support this highly targeted communication strategy, our GDPR friendly data is designed to work alongside data agency contact lists to enhance the effectiveness of selection rules. It has a near 100% coverage across the UK and is highly localised, being postcode tagged. It is also specifically related to COVID-19 risks and outcomes, making it a unique “added- value” resource that can be used by communication professionals within existing CDM operations quickly.

By way of example, if we compare the profile for the three Welsh constituencies to the three for Leicester we see contrasting profiles.

The Welsh profile is concentrated in Zone D where age risks are high. The message that needs to be communicated here is for the older, at-risk to keep self-isolating so that they are not exposed. Extra support to this vulnerable group and their carers where relevant should be the focus whilst the infection rate is brought down. If visitors / tourists are the reason infection rates are staying high, then additional steps to control this source should be considered to ensure they are kept well away from vulnerable citizens.

In Leicester the risk profile is concentrated near Zones B and C (Overcrowding and Health Risk respectively). If the infections are remaining high in these types of neighbourhood it indicates that there is a need for a concerted set of actions to stop infection rates quickly getting out of control. Individuals with additional risks in these neighbourhoods (e.g. those with obesity or other health issues, or high risks related to ethnicity) should be of particular concern, especially if exposure through other members of the household is likely to be a factor.





Case Study 2

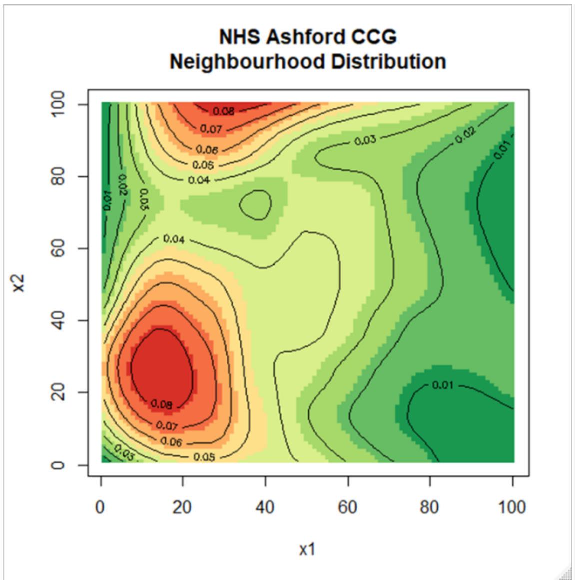
NHS Practice Manager

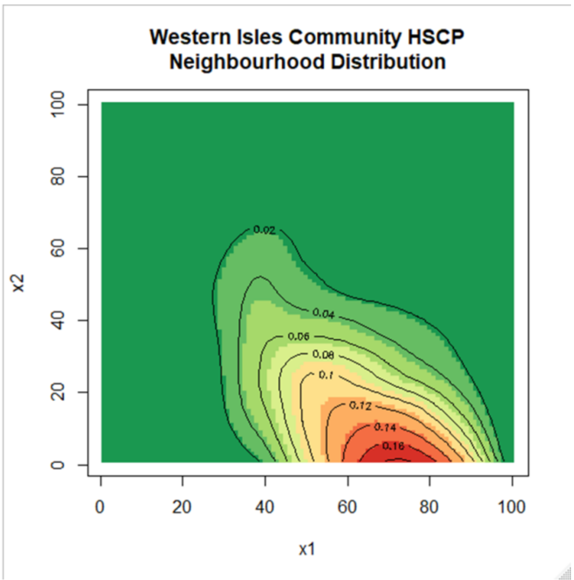
Burning Issue: Will we have sufficient capacity to cope with a second wave?

Suggested Approach. Compare an Organisation's "patch" to other NHS catchments and identify its risk profile is relative to others. Use this insight to compare each NHS organisation with its peers to see how well it is able to balance potential demand with available resources. Are individual organisations within the NHS appropriately resourced or not on this basis? Also use this insight to get effective conversations going across the wider organisation and into the local community. What is working well for each local organisation? What can we learn from other NHS organisations and form the community to address the areas of highest risk going forward?

Why use More Metrics Data to support this activity? Our data provides an "out of the box" set of data that is specifically designed to assess COVID risk at a very local level across all parts of the UK on a robust and consistent basis. Our postcode tagged data can therefore be diced and sliced to generate benchmarking reports at whatever geographical level is needed. It is also GDPR friendly because it only uses aggregate, publicly available data so there are no restrictions on sharing this analysis outside of a local organisation. This also means there is no need to get hold of data from peers to do comparisons as it is already available from us for the whole of the UK. This means organisations can move fast, enabling them to be proactive in contacting those NHS colleagues elsewhere that they can learn most from.

To give one illustration of these ideas, we show below how two CCGs that are at opposite ends of the "Time Adjusted" infection rate spectrum compare using our COVID Risk Map and Wave 2 risk grid. NHS Ashford has a very high "Time Adjusted" infection rate (23%) and Western Isles a very low rate (3%) as at 21st June 2020





NHS Ashford CCG has a neighbourhood distribution that spans the age range from young to middle age with fewer neighbourhoods in the very oldest risk group. Residents are in socially and economically active parts of the COVID Risk Map, with some neighbourhoods concentrated in the early and high infection risk part of the COVID Map (Top, Left / Middle).

By contrast, Western Isles Community HSCP has neighbourhoods that are over-indexed in the highest age groups (Bottom, Middle / Right). We can see that there are few if any high-infection-risk, younger neighbourhoods in the Western isles. This limits the rate of community transmission to the highly vulnerable older neighbourhoods which greatly lowers the year to date infection rates here.

This results in a different assessment of the Wave 2 risk for these two areas

Highest Time Adjusted Infection Rate: NHS Ashford

Potential Risk Rank ("Time Adjusted")				
	High	9%	18%	46%
	Medium	3%	3%	8%
	Low	1%	4%	6%
More Metrics Risk Matrix		Low	Medium	High
		Actual Risk Rank ("As Is")		

Lowest Time Adjusted Infection Rate: Western Isles Community HSCP

Potential Risk Rank ("Time Adjusted")				
	High	0%	0%	0%
	Medium	3%	3%	0%
	Low	94%	2%	0%
More Metrics Risk Matrix		Low	Medium	High
		Actual Risk Rank ("As Is")		

The NHS Ashford Risk Matrix (Left) has roughly half its Output Areas in the High / High cell. For neighbourhoods in this Risk Cell we estimate an average cumulative infection rate of 36% as at 21st June 2020. This is starkly different to Western Isles (Right) which has 94% of neighbourhoods in the Low / Low Cell with an estimated average infection rate of 1.5% as at 21st June 2020.

Recall from above that the highest Wave 2 Risk is expected to be in the top left cell (High Potential Infection Rate / Low Actual Infection Rate). In Ashford's case we estimate that 9% of their neighbourhoods are in this cell with an estimated infection rate as at 21st June of 16%, which is less than half that in their High / High cell. Armed with the information on which postcodes fall into which cell on our matrix, Ashford can approach PHE to get confirmation on likely infection rates from a national analysis of their antibody test statistics. Using the PHE results, average infection rates for all nine cells of the grid for England as a whole can be calculated from the list of More Metrics postcodes that fall into each cell.

Overlaying an age band split on top of the grid, should confirm whether heterogeneous infection rates locally mean that the Herd Immunity Threshold (HIT) is being achieved at least in some younger sub populations of the Ashford CCG and not being achieved for other neighbourhoods where Wave 2 risks remain very high[10]

In the case of Western Isles, it is apparent that infection rates are low everywhere. HIT will not apply, and the requirement going forward is to maintain social distancing of their vulnerable, older residents over the long term until a vaccine is available. This may require the NHS to request additional actions as lockdown is relaxed, particularly if "tourist incomers" are identified as their greatest risk to future resources. Taking practical steps such as the use of ribbons or wrist bands to help residents communicate their risk status to strangers may be of particular value in this situation[11].

[10] See the paper by Gomes et al for a detailed analysis of the heterogeneous infection rate issue https://www.researchgate.net/publication/341108416_Individual_variation_in_susceptibility_or_exposure_to_SARS-CoV-2_lowers_the_herd_immunity_threshold

[11] https://www.dailymail.co.uk/news/article-8445717/Sage-adviser-suggests-elderly-people-wear-ribbons-indicate-social-distancing.html

Case Study 3	
Sales and Marketing Director of a Retail Business	**Burning Issue:** How do we adapt to rapidly changing B2C buying behaviour and predict where it might be heading so that we can re-build our balance sheets? **Some suggestions:** Applying the Donald Rumsfeld approach to problem structuring in this situation is probably a good place to start[12]. For example, existing segmentation and time series analysis can identify how a lot of the "known knowns" have played out as we have entered into lockdown and are now coming out. This would include charting the acceleration in well-established trends of people purchasing more on-line split by customer segment and prior channel. Trending this forward under a range of different scenarios can help get a handle on the known unknowns such as how quickly (if ever) will people go back to their old ways of shopping when outlets re-open? If this happens slowly or never, what does this mean for the bricks and mortar estate and the staff that work there? By definition, the unknown unknowns can't be analysed until they become known unknowns or known knowns, so the challenge is to set up an analytical approach that is designed to spot emerging issues quickly and to respond optimally to events as they occur. One example where this is particularly important is to be alert to competitor responses where it is difficult to know how things might play out and where scenario planning can be unwieldy when things are so uncertain because there are just too many factors to consider right now. **Why use More Metrics Data to support this activity?** Speed of response and proactivity is enabled by being prepared. This means having relevant analysis at your fingertips that is forward looking and updated regularly. Our data is comprehensive in scope covering 20 measures of COVID Risk with millions of point estimates tagged by postcode and time point, updated weekly. This is tailor made for integrating into all existing customer segmentation, share of wallet and SKU time series models that can therefore be updated and re-purposed quickly to deal with emerging issues. In addition we have "off the shelf" companion datasets that can fill other gaps such as our geo money and geo lifestyle series. All of our data is modelled from open-source data that ensures our data is GDPR friendly and provides full UK coverage. As required we are happy to provide bespoke datasets that fill any other gaps identified. One area that we think may be of particular value is in identifying those consumers most financially impacted by COVID 19 as the support offered by governments across the UK is gradually withdrawn. We have established techniques for imputing local estimates of census micro data that can be particularly helpful for this type of analysis. We have already applied this approach for one of our risk rank measures (employment risk) and can update this as required to reflect recent events, working with our established data partners as appropriate to ensure it is kept fully up-to-date.

[12] "There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know" Donald Rumsfeld quote.

Concluding Remarks	We believe that our COVID-19 Risk datasets provide a unique set of measures for anyone involved in managing the COVID epidemic in whatever role. Our free to use datasets are aimed at those responsible for charting the progress of the infection for their organisation and provide a useful set of geographical cuts of the data that are designed to make this as straightforward as possible without skimping on useful detail. For power users who are building models or who have access to a wide range of other data sources that may include data on individuals, our detailed datasets at Output Area and LSOA provide complementary views on risk that we believe are unique. Our postcode tagged dataset provide millions of infection rate estimates and other data designed to support classification of risks in many different ways at a very local level. Our data should be of particular value where analytical teams are dealing with significant data gaps and where there is a need to make sense of source data that appears to be predictive but may not generalise well enough to give robust answers to decision makers about the pros and cons of different courses of action.
Obtaining the Data	Our products are available directly from More Metrics or through one of our partnerships with leading data agencies. Follow the links on our website to get access to our data. The Ward, Parliamentary Constituency, Local Authority and CCG datasets can be downloaded in a single excel workbook from the More Metrics website by registering your details. Alternatively, individual files of data can be obtained from our data agency partners. These data distributors can also supply more geographical detailed datasets on a commercial basis if required. Special rates are available for those users who can demonstrate that their use of our detailed data is only for non-commercial reasons that support the public good.

Data acknowledge-ments and attributions	The COVID-19 dataset contains data from other sources which have their own copyright notice as follows: Contains OS data © Crown copyright and database rights 2020 Contains Royal Mail data © Royal Mail copyright and database rights 2020 Contains National Statistics data © Crown copyright and database rights 2020 Contains Public Health England Data © Crown copyright and database rights 2020 More Metrics retains copyright to the rest of the COVID-19 data derived from open source data © More Metrics copyright and database rights 2020. Our main source of data for models is 2011 census data supplemented by a wide-range of more up to date data provided by National Records of Scotland (Crown Copyright, OGL), Northern Ireland Statistics and Research Agency (Crown Copyright, OGL), Office of National Statistics (Crown Copyright, OGL). Output Area mappings to other Geographies are taken from the ONSPD / NSPL files and other lookup files regularly published by ONS. These files contain National Statistics data © Crown copyright and database right 2020 https://www.ons.gov.uk/methodology/geography/geographicalproducts/postcodeproducts Other data sources used by us in feeder models for the COVID-19 dataset are Crown Copyright and used under Open Government Licence v.3.0, as follows: Inheritance Tax model uses data published by HMRC Earned Income model uses Annual Survey of Hours and Earnings (ASHE) data published by ONS Obesity and Smoker models use data published by Public Health England (PHE) and ONS Engagement risk models contain Parliamentary information licensed under the Open Parliament Licence v3.0 https://www.parliament.uk/site-information/copyright/open-parliament-licence/ The data used for the response curve analysis was downloaded from the Our World in Data GitHub site. More Metrics is very grateful to the OWID authors for making this resource available and we confirm that our use of their data is for non-commercial purposes as a standalone piece of work that is not included in any of our commercial datasets. The R-code we have developed for the response curve analysis using the OWID data is available on request for anyone interested. OWID data and license details can be found on their website and the salient details are copied below https://github.com/owid/covid-19-data/tree/master/public/data *"Licence* *All of Our World in Data is completely open access and all work is licensed under the Creative Commons BY license. You have the permission to use, distribute, and reproduce in any medium, provided the source and authors are credited.* *Authors* *This data has been collected, aggregated, and documented by Diana Beltekian, Daniel Gavrilov, Charlie Giattino, Joe Hasell, Bobbie Macdonald, Edouard Mathieu, Esteban Ortiz-Ospina, Hannah Ritchie, Max Roser."* As at 22/06/2020

Appendix 1 — Selected conversion tables used to calculate infection rates

Age	Average Risk index (Travel, Room, Resident)	Mingler Proportion	Enforced-Mingler Proportion	Non-Mingler Proportion	More Metrics Infection Rate Index (Age 50 = 1)	Symptomatic Covid Infection Rate Index* (Age 50 = 1)	More Metrics Death Rate Index (IFR basis)** (Age 50 = 1)	More Metrics Death Rate Index (CFR basis)** (Age 50 = 1)	ONS Death Rate Index (CFR basis)** (Age 50 = 1)
Age 18	54	0.5	0.01	0.49	1.48	0.55	0.02	0.03	0.02
Age 19	53	0.5	0.01	0.49	1.47	0.55	0.02	0.03	0.02
Age 20	53	0.5	0.01	0.49	1.46	1.02	0.02	0.03	0.03
Age 21	54	0.49	0.01	0.5	1.45	1.02	0.02	0.03	0.03
Age 22	54	0.49	0.01	0.5	1.44	1.02	0.03	0.04	0.03
Age 23	54	0.48	0.01	0.51	1.42	1.02	0.03	0.04	0.03
Age 24	54	0.48	0.01	0.51	1.41	1.02	0.03	0.04	0.03
Age 25	54	0.47	0.01	0.52	1.4	1.02	0.04	0.06	0.06
Age 26	55	0.47	0.01	0.52	1.38	1.02	0.04	0.06	0.06
Age 27	55	0.46	0.01	0.53	1.36	1.02	0.05	0.07	0.06
Age 28	55	0.45	0.01	0.54	1.35	1.02	0.06	0.08	0.06
Age 29	55	0.44	0.01	0.55	1.33	1.02	0.06	0.08	0.06
Age 30	55	0.43	0.01	0.56	1.31	0.91	0.07	0.09	0.1
Age 31	55	0.43	0.01	0.56	1.29	0.91	0.08	0.1	0.1
Age 32	55	0.42	0.01	0.57	1.26	0.91	0.09	0.11	0.1
Age 33	55	0.41	0.01	0.58	1.25	0.91	0.11	0.14	0.1
Age 34	55	0.4	0.01	0.59	1.23	0.91	0.12	0.15	0.1
Age 35	54	0.39	0.01	0.6	1.21	0.91	0.14	0.17	0.16
Age 36	54	0.38	0.01	0.61	1.19	0.91	0.16	0.19	0.16
Age 37	54	0.37	0.01	0.62	1.17	0.91	0.18	0.21	0.16
Age 38	53	0.37	0.01	0.62	1.15	0.91	0.2	0.23	0.16
Age 39	53	0.36	0.01	0.63	1.13	0.91	0.23	0.26	0.16
Age 40	53	0.35	0.01	0.64	1.12	0.96	0.27	0.3	0.34
Age 41	52	0.35	0.01	0.64	1.1	0.96	0.3	0.33	0.34
Age 42	52	0.34	0.01	0.65	1.08	0.96	0.35	0.38	0.34
Age 43	51	0.33	0.01	0.66	1.07	0.96	0.4	0.43	0.34
Age 44	51	0.33	0.01	0.66	1.06	0.96	0.45	0.48	0.34
Age 45	51	0.32	0.01	0.67	1.04	0.96	0.52	0.54	0.56
Age 46	51	0.32	0.01	0.67	1.03	0.96	0.59	0.61	0.56
Age 47	50	0.32	0.01	0.67	1.02	0.96	0.68	0.69	0.56
Age 48	50	0.31	0.01	0.68	1.02	0.96	0.77	0.79	0.56
Age 49	50	0.31	0.01	0.68	1.01	0.96	0.88	0.89	0.56
Age 50	49	0.31	0.01	0.68	1	1	1	1	1
Age 51	49	0.3	0.01	0.69	0.99	1	1.13	1.12	1
Age 52	49	0.3	0.01	0.69	0.98	1	1.28	1.25	1
Age 53	49	0.3	0.01	0.69	0.98	1	1.44	1.41	1
Age 54	48	0.3	0.01	0.69	0.98	1	1.62	1.59	1
Age 55	48	0.3	0.01	0.69	0.98	1	1.81	1.77	1.85
Age 56	47	0.29	0.01	0.7	0.97	1	2.02	1.96	1.85
Age 57	47	0.29	0.01	0.7	0.97	1	2.24	2.17	1.85
Age 58	47	0.29	0.01	0.7	0.97	1	2.48	2.41	1.85
Age 59	46	0.29	0.01	0.7	0.97	1	2.73	2.65	1.85
Age 60	46	0.29	0.01	0.7	0.96	0.62	3	2.88	3.09
Age 61	45	0.29	0.01	0.7	0.96	0.62	3.3	3.17	3.09
Age 62	45	0.29	0.01	0.7	0.96	0.62	3.62	3.48	3.09
Age 63	44	0.29	0.01	0.7	0.96	0.62	3.98	3.82	3.09
Age 64	44	0.29	0.01	0.7	0.95	0.62	4.37	4.15	3.09
Age 65	44	0.29	0.01	0.7	0.95	0.62	4.82	4.58	4.48
Age 66	44	0.29	0.01	0.7	0.95	0.62	5.32	5.05	4.48
Age 67	43	0.27	0.03	0.7	0.95	0.62	5.9	5.61	4.48
Age 68	43	0.27	0.03	0.7	0.94	0.62	6.56	6.17	4.48
Age 69	43	0.27	0.03	0.7	0.94	0.62	7.32	6.88	4.48
Age 70	43	0.27	0.03	0.7	0.94	0.43	8.19	7.7	7.74
Age 71	43	0.27	0.03	0.7	0.93	0.43	9.21	8.57	7.74
Age 72	43	0.27	0.03	0.7	0.93	0.43	10.4	9.67	7.74
Age 73	43	0.27	0.03	0.7	0.93	0.43	11.78	10.96	7.74
Age 74	43	0.27	0.03	0.7	0.92	0.43	13.4	12.33	7.74
Age 75	42	0.27	0.03	0.7	0.91	0.43	15.32	13.94	16.61
Age 76	42	0.27	0.03	0.7	0.89	0.43	17.58	15.65	16.61
Age 77	42	0.25	0.05	0.7	0.88	0.43	20.27	17.84	16.61
Age 78	42	0.25	0.05	0.7	0.86	0.43	23.49	20.2	16.61
Age 79	42	0.25	0.05	0.7	0.84	0.43	27.38	23	16.61
Age 80	42	0.23	0.07	0.7	0.82	0.45	32.09	26.31	32.11
Age 81	41	0.21	0.09	0.7	0.79	0.45	37.83	29.89	32.11
Age 82	41	0.21	0.09	0.7	0.75	0.45	44.89	33.67	32.11
Age 83	41	0.19	0.11	0.7	0.72	0.45	53.58	38.58	32.11
Age 84	41	0.15	0.15	0.7	0.68	0.45	64.34	43.75	32.11
Age 85	41	0.13	0.17	0.7	0.64	0.45	77.65	49.7	57.62
Age 86	40	0.09	0.21	0.7	0.6	0.45	94.13	56.48	57.62
Age 87	40	0.07	0.23	0.7	0.55	0.45	114.49	62.97	57.62
Age 88	40	0.03	0.27	0.7	0.51	0.45	139.58	71.19	57.62
Age 89	40	0.01	0.29	0.7	0.47	0.45	170.41	80.09	57.62
Age 90 and Over	40	0	0.41	0.59	0.43	0.45	208.17	89.51	109.6

*Source figures are taken from data published by King's college London using the results from the Covid Sympton Study App developed by by health science company ZOE
** IFR basis assumes everyone is infected (the Imperial College do nothing scenario). CFR basis reflects the observed death rates for known cases
CFR Source figures are estimated from Professor David Spiegelhalter's paper which used ONS published death data for England and Wales up to May 29th