MODELLING THE SPREAD OF COVID-19 USING NON-STANDARD MEASURES OF POPULATION DENSITY

Professor Oliver Johnson
Professor of Information Theory, School of Mathematics, University of Bristol.

Since the start of the coronavirus pandemic, there has been considerable interest in comparing the response of different countries, using numbers of deaths or positive tests to measure the effectiveness of their policy and healthcare response. While such calculations are tempting, they need to be performed carefully, and with certain caveats in mind.

First, there are issues of timing and reporting. Not all countries are at a comparable stage of their own epidemic, and standards vary hugely in terms of how and when deaths are reported, and which individuals are included in national figures. For this reason, it is premature to compare countries too early in the pandemic: in the absence of a vaccine, it may well be that final death tolls will be fairly similar overall when viewed as a proportion of the population.

Second, and more importantly, not all countries are alike. For example, it was relatively easy for New Zealand to follow a strategy of isolation and elimination, due to its geographical isolation. It is far from clear that, given the presence of a major international hub airport and large numbers of international visitors including students, this would have been a realistic option for the UK.

Further, there are other factors that affect the severity of the virus which mean that not all countries should expect the same number of deaths. For example, fatality rates are significantly higher among the old, so one might expect that countries with a higher average age or a larger proportion of over 70s would expect to see more COVID casualties. This might suggest why fatality rates are so much worse in South American than in Africa at the time of writing. Similarly, levels of obesity or diabetes are also associated with worse healthcare outcomes for coronavirus, so should be taken account of when assessing international casualty figures.

One particular demographic factor that appears natural in this context is population density. We will refer to the usual calculation of total people per square kilometre in a region as the *standard population density*, to distinguish it from another measure we discuss later. Having seen the severity of the outbreak in New York City and other densely populated regions, it is natural to hypothesise that the higher the standard population density, the faster the virus should spread. In more crowded areas, it seems inevitable that

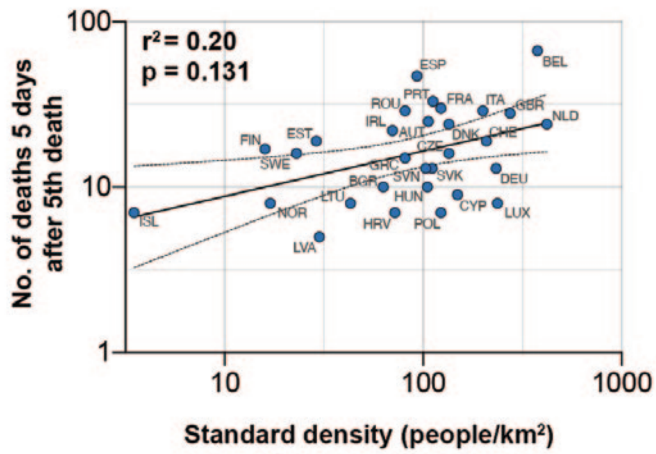

Figure 1: relationship between standard population density and rate of spread of COVID-19 for a number of European countries.

people will have more close contact with others, so an infected individual will transmit the coronavirus to more people, and the outbreak will grow at a faster rate.

This is a hypothesis that is relatively easy to test. National population data is readily available, and a number of websites aggregate coronavirus casualties. Hence, we can calculate the standard population density for each country, and plot it against the rate of spread early in the epidemic. For example, we might look at how many deaths took place in the 5 days following the casualty figures reaching 5 deaths per day.

As we see from Figure 1, there is surprisingly little correlation between standard population density and the rate of spread, when comparing European countries. In general, countries with larger standard population densities typically have faster spread of the virus, but there are many exceptions, and the trend is not statistically significant. We can perhaps understand why by looking at Spain. Although the virus spread very fast there, taken as a whole the country has a very low standard population density of 93 people per square kilometre. However, this low figure reflects the fact that the country contains many empty regions where nobody at all lives, as well as many of the highest density neighbourhoods in Europe, in Barcelona and Madrid.

We can see a similar phenomenon in New York State, another hotspot for the virus. Again, the standard population density is very low (163 people per square kilometre), but this is made up of a combination of relatively empty areas of land upstate and the extremely high density areas of New York City itself.

This helps us to understand why the standard population density is not the right measure in this context. One way to see this is to understand what it means: the standard population density tells us "how many people we expect to be living next to a randomly chosen point in the country?". In that sense, the large empty spaces of upstate New York count for more than the relatively small area of Manhattan. However, we need to think from a different point of view: that of the virus.

The virus does not pick a random point in space: it effectively picks a random person. For example, we can imagine that an initial outbreak will be seeded by an international traveller arriving from outside the region. In that case, the right question to ask is "how many people we expect to be living next to a randomly chosen person?". It is perhaps not obvious that this is a different question to the one above. However, it is clear that sampling in this way will weight Manhattan much more highly in the calculation, since a randomly chosen person is more likely to live there. This leads us to use the *quadratic population-weighted density* as an alternative to the standard population density. Thanks to data provided by the WorldPop project in Southampton, we have access to population data on the scale of a square kilometre grid, so can calculate this relatively easily.

It turns out that Spain has a particularly high value of this population-weighted density (3273 people per square kilometre), as does New York State (6163 people per square kilometre). We can plot this population-weighted density against the rate of spread. We find that the population-weighted density does a better job of explaining the rate of spread than the standard density when comparing European countries (see Figure 2), explaining roughly half the variation observed between countries.

Making this comparison allows us to see which countries are performing particularly well or badly, through having a faster or slower rate of spread respectively than their population-weighted density suggests. In particular, given the large amount of media interest comparing the epidemic in the UK and Germany, it is interesting to notice that neither country stands out as an outlier in that sense. Indeed, it may be somewhat surprising to find that Germany has a population-weighted density of only 885 people per square kilometre – lower than Sweden or Ireland – reflecting the fact that its population is fairly evenly distributed across the country, making the virus slower to spread. In fact, the country which stands out for having a slower spread than expected is not Germany but Greece, which locked down early to very positive effect.

Of course, the population density is not the only factor that explains the spread of the virus, but we argue it must be taken into account when comparing outcomes between different countries. However, we emphasise that any such comparisons must be performed carefully and rigorously, and only at the end of the pandemic.

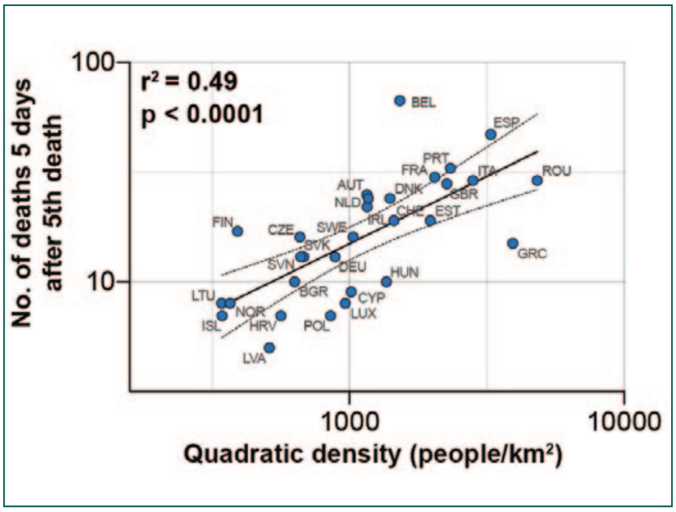

Figure 2: relationship between population-weighted density and rate of spread of COVID-19 for a number of European countries.

Reference:

Preprint: Garland, P., Babbitt, D., Bondarenko, M., Sorichetta, A., Tatem A. and Johnson, O. Lived population density and the spread of COVID-19 https://arxiv.org/pdf/2005.01167.pdf (2020) ☐