

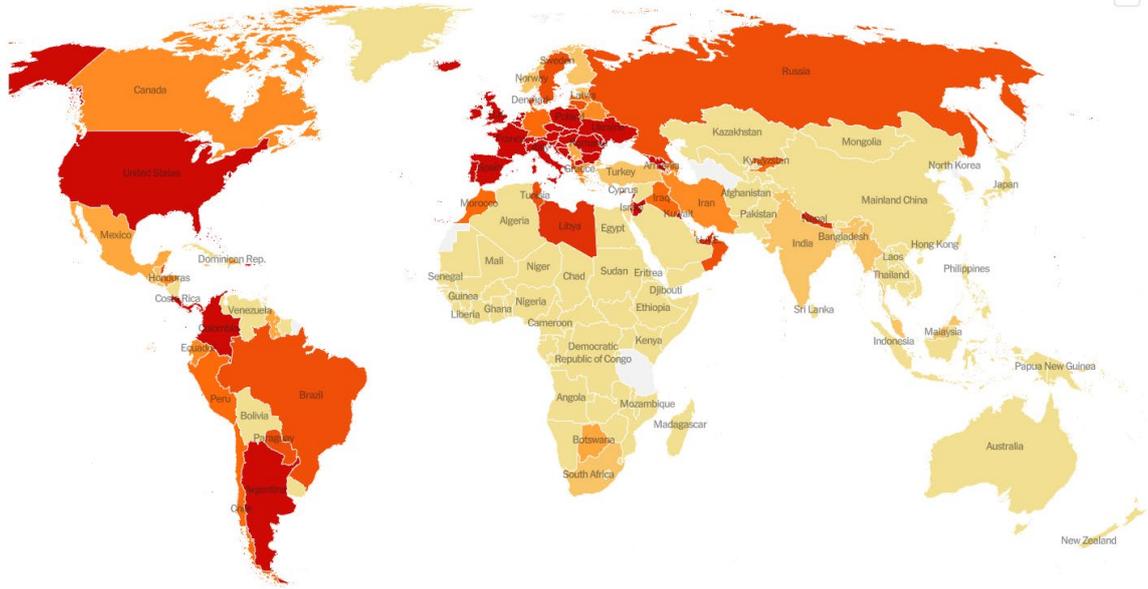
## How Overdispersion Drives the Spread of the SARS-CoV-2 Pandemic

Paul Somerville, Risk Frontiers

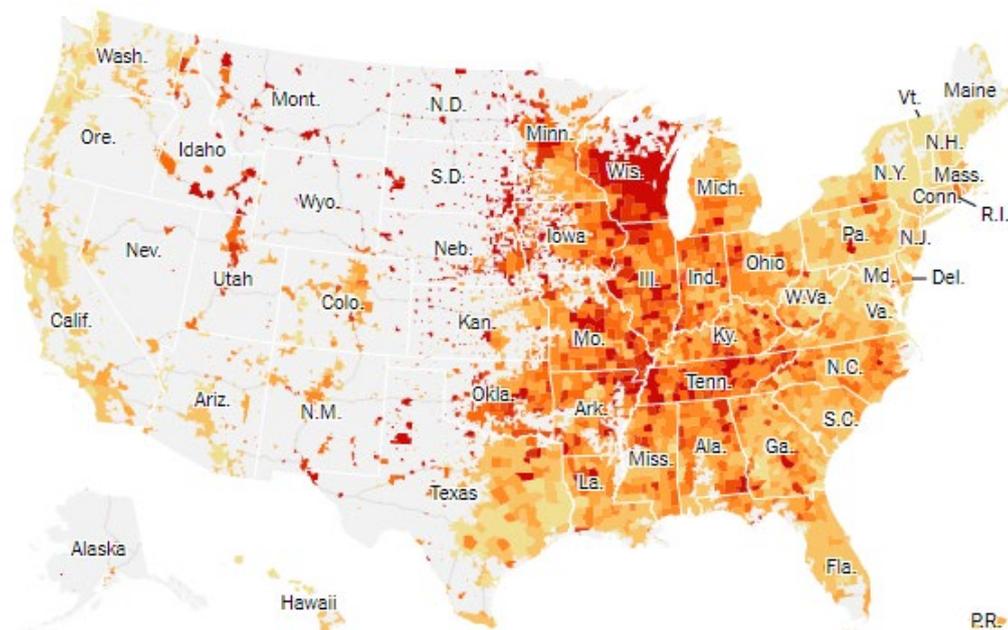
Even after months of extensive research by the global scientific community, many questions about the spread of the SARS-Covid-2 pandemic remain unanswered. Widespread expectations of catastrophic outbreaks in China, South Korea and Japan were not realized. In the early months of 2020, a few cities such as New York accounted for a substantial portion of global deaths, while many others with similar population density, weather, age distribution, and travel patterns were spared. There was an enormous death toll in northern Italy, but not the rest of the country. In Guayaquil, Ecuador, so many people died so quickly in April that bodies were abandoned in streets. A snapshot of this heterogeneity in the average daily number of cases per 100,000 people by country and within the United States in the week of October 18-24 is shown in Figures 1 and 2 respectively.

Population-level analyses often use average quantities to describe heterogeneous systems, particularly when variation does not arise from identifiable groups. A prominent example, central to the understanding of epidemic spread, is the basic reproductive number,  $R_0$ , which is defined as the mean number of infections caused by an infected individual in a susceptible population. Most of the discussion about the spread of SARS-CoV-2 has concentrated on  $R_0$  which, without social distancing, is about 3. In real life, however, some people infect many others and others do not spread the disease at all, so the most common individual number is zero. Hence population estimates of  $R_0$  can obscure considerable individual variation in infectiousness, as highlighted during the global emergence of severe acute respiratory syndrome (SARS) and SARS-CoV-2 by numerous 'superspreading events' in which certain individuals infected unusually large numbers of secondary cases.

Lloyd-Smith et al. (2005) introduced the 'individual reproductive number',  $v$ , as a random variable representing the expected number of secondary cases caused by a particular infected individual. Values for  $v$  are drawn from a continuous probability distribution with population mean  $R_0$  that encodes all variation in infectious histories of individuals, including properties of the host and pathogen and environmental circumstances. In this framework, superspreading events are not exceptional events, but important realizations from the right-hand tail of a distribution of  $v$ . Stochastic effects in transmission are modelled using a Poisson process, so that the number of secondary infections caused by each case,  $Z$ , is described by an 'offspring distribution'  $\Pr(Z = k)$  where  $Z \sim \text{Poisson}(v)$ . In their preferred model, Lloyd-Smith et al. (2005) let  $v$  be gamma-distributed with mean  $R_0$  and dispersion parameter  $k$ , yielding  $Z \sim \text{negative binomial}(R_0, k)$ . The negative binomial model includes the conventional Poisson ( $k \rightarrow \infty$ ) and geometric ( $k = 1$ ) models as special cases. It has variance  $R_0(1 + R_0/k)$ , so smaller values of  $k$  indicate greater heterogeneity. The parameter  $k$  is referred to as the overdispersion, and represents the random variability in the mean of the Poisson distribution. An analogous departure from a conventional Poisson model was encountered in our briefing on the temporal clustering of large earthquakes (Briefing note 412).



**Figure 1.** Average daily cases per 100,000 people by country in the week of October 18-24, 2020. Source: New York Times. <https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html#cases>



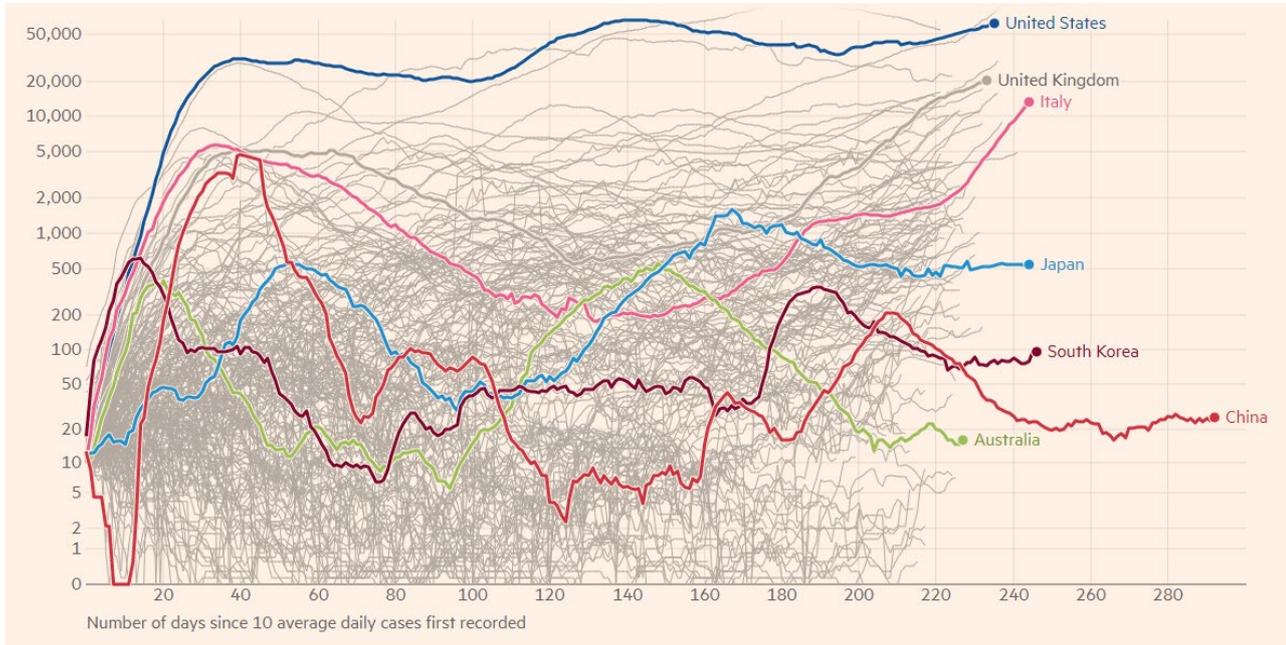
**Figure 2.** Average daily cases per 100,000 people in the United States in the week of October 18-24, 2020. Source: New York Times <https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html#cases>

The role of overdispersion in the spread of SARS-Covid2 has been described by Kupferschmidt (2020) and Tufekci (2020). Lloyd-Smith et al. (2005) estimated that SARS - in which superspreading played a major role - had a  $k$  of 0.16. The estimated  $k$  for MERS, which emerged in 2012, is about 0.25. In the flu pandemic of 1918, in contrast, the value was about one, indicating that spatiotemporal clusters played less of a role. The most recent estimate of  $k$  for SARS-CoV-2 is about 0.1 (Endo et al., 2020), indicating a higher spatiotemporal clustering rate than for all the other pandemics, with 10% of the cases leading to about 80% of the spread. This kind of behavior, alternating between being super infectious and fairly noninfectious, is what  $k$  captures and what focusing solely on  $R_0$  hides. This has presented a large challenge, especially for health authorities in Western societies, where the approach to controlling the pandemic has been modeled on the flu.

Disease patterns can be thought of as having deterministic or stochastic trends. In the deterministic case, an outbreak's distribution is more linear and predictable, while in the stochastic case, randomness plays a much larger role and predictions are hard, if not impossible, to make. In deterministic trajectories, what happened yesterday is expected to give a good estimate of what to expect tomorrow. Diseases like the flu are practically deterministic and adequately represented by  $R_0$ , and they are nearly impossible to stop until there is a vaccine. Stochastic phenomena, however, do not operate that predictably, and random variations can rapidly tip conditions from one state to another.

The highly skewed, imbalanced distribution of  $k$  in stochastically spreading cases like SARS-CoV-2 means that an early run of bad luck with a few super-spreading events, or clusters, can produce dramatically different outcomes even for otherwise similar countries, as occurred in Italy and South Korea (Figure 3). Scientists who have looked globally at known early-introduction events, in which an infected person comes into a country, found that in some places, such imported cases led to no deaths or known infections, while in others, they sparked sizable outbreaks. This could explain some puzzling aspects of this pandemic, including why the virus did not rapidly spread around the world sooner after it emerged in China, and why some very early cases elsewhere - such as one in France in late December 2019, apparently failed to ignite a wider outbreak. If  $k$  is really 0.1, then most chains of infection die out by themselves and SARS-CoV-2 needs to be introduced undetected into a new country at least four times to have an even chance of establishing itself. Most of the cases that left China simply fizzled out. Using genomic analysis, researchers in New Zealand looked at more than half the confirmed cases in the country and found 277 separate introductions in the early months, but that only 19 percent of introductions led to more than one additional case. This may even be true in congregate living spaces, such as nursing homes, and multiple introductions may be necessary before an outbreak takes off.

Meanwhile, in Daegu, South Korea, just one woman, dubbed Patient 31, generated more than 5,000 known cases in a megachurch cluster. Nevertheless, overdispersion is also a cause for hope, as demonstrated by South Korea's aggressive and successful response to that outbreak with a massive testing, tracing, and isolating regime. Since then, South Korea has also been practicing sustained vigilance, and has demonstrated the importance of backward tracing. When a series of clusters linked to nightclubs broke out in Seoul recently, health authorities aggressively traced and tested tens of thousands of people linked to the venues.



**Figure 3.** Seven-day rolling average of new cases in selected countries, by number of days since 10 average daily cases first recorded, October 26, 2020. Source: Financial Times.

Australia targeted mass suppression of the virus with a focus on clusters, and generally succeeded in this approach, with the temporary exception of widespread community transmission in Melbourne due to failure of quarantine measures. Australia invested in widespread testing early on and used mechanisms to monitor contact tracing.

One of the most interesting cases has been Japan, a country that was hit early on, like Australia, by passengers from a cruise ship. Japan followed what appeared to be an unconventional model, not deploying mass testing and never fully shutting down. By the end of March, influential economists were publishing reports with dire warnings, predicting overloads in the hospital system and huge spikes in deaths. The predicted catastrophe never came to be, however, and although it faced some future waves, there was never a large spike in deaths despite its aging population, uninterrupted use of mass transportation, dense cities, and lack of a formal lockdown.

In the beginning, Japan was no better situated than other countries such as the United States and those in Western Europe that now have much worse outcomes. Like them, Japan did not initially have the capacity to do widespread testing. Nor could Japan impose a full lockdown or strict stay-at-home orders, because even if that had been desirable, it would not have been legally possible. Instead, the Japanese specialists had noticed the overdispersion characteristics of COVID-19 as early as February, and so they created a strategy focusing mostly on cluster-busting, which tries to prevent one cluster from igniting another. This entailed aggressive backward tracing to uncover clusters. Japan also focused on ventilation and counseling its population to avoid places where the three C's come together - crowds in closed spaces in close contact. Cultural factors including wearing masks in public especially during the flu season, and consideration of the welfare of others.

This strategy is in contrast with the Western response, which tries to eliminate the disease case by case, when that is not necessarily the main way it spreads. Japan did get its cases down and kept up its vigilance. When the government started noticing an uptick in community cases, it initiated a state of

emergency in April and tried hard to incentivize the kinds of businesses that could lead to super-spreading events, such as theatres, music venues, and sports stadiums, to close down temporarily. Now schools are back in session in person, and even stadiums are open, but without chanting.

Countries that have ignored super-spreading have risked getting the worst of both worlds: burdensome restrictions that fail to achieve substantial mitigation. The UK's recent decision to limit outdoor gatherings to six people while allowing pubs and bars to remain open is just one of many such examples. Many studies have shown that super-spreading clusters of COVID-19 almost overwhelmingly occur in poorly ventilated, indoor environments where many people congregate over time - weddings, churches, choirs, gyms, funerals, restaurants, and such - especially when there is loud talking or singing without masks. For super-spreading events to occur, multiple things have to be happening at the same time, and the risk is not the same in every setting and activity.

If public health workers know where clusters are likely to happen, they can try to prevent them and avoid shutting down broad swaths of society. Lockdowns are a blunt tool, because they concede that not enough is known about where transmission is happening to be able to target it, so everything is indiscriminately targeted. Unfortunately, studying large COVID-19 clusters is harder than it seems. Many countries have not collected the kind of detailed contact tracing data needed, and the lockdowns have been so effective that they also robbed researchers of a chance to study superspreading events. Before the lockdowns, there was probably a 2-week window of opportunity when a lot of these data could have been collected.

## References

Endo A, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott S *et al.* Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China [version 3; peer review: 2 approved]. *Wellcome Open Res* 2020, 5:67

(<https://doi.org/10.12688/wellcomeopenres.15842.3>)

Financial Times (2020). <https://www.ft.com/coronavirus-latest>

Kupferschmidt, Kai (2020). Why do some COVID-19 patients infect many others, whereas most don't spread the virus at all? *Science Magazine*, May 19, 2020.

<https://www.sciencemag.org/news/2020/05/why-do-some-covid-19-patients-infect-many-others-whereas-most-don-t-spread-virus-all>

Lloyd-Smith, J., Schreiber, S., Kopp, P. *et al.* Superspreading and the effect of individual variation on disease emergence. *Nature* 438, 355–359 (2005). <https://doi.org/10.1038/nature04153>

[Risk Frontiers \(2020\). Devil's Staircase of Earthquake Occurrence: Implications for Seismic Hazard in Australia and New Zealand. Briefing note 412, April 2020.](#)

Tufekci, Zeynep (2020). This Overlooked Variable Is the Key to the Pandemic It's not R. *The Atlantic*, September 30, 2020

[https://www.theatlantic.com/health/archive/2020/09/k-overlooked-variable-driving-pandemic/616548/?utm\\_source=newsletter&utm\\_medium=email&utm\\_campaign=atlantic-daily-newsletter&utm\\_content=20201020&silverid-ref=MzEwMTkwMTM3ODg3So](https://www.theatlantic.com/health/archive/2020/09/k-overlooked-variable-driving-pandemic/616548/?utm_source=newsletter&utm_medium=email&utm_campaign=atlantic-daily-newsletter&utm_content=20201020&silverid-ref=MzEwMTkwMTM3ODg3So)