ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

12 Errol Street
London EC1Y 8LX

020 7638 8998
rss@rss.org.uk
www.rss.org.uk

William Wragg MP
Chair, Public Administration and Constitutional Affairs Committee
House of Commons
London
SW1A 0AA

By email

08/12/2020

Dear Mr Wragg,

**Data transparency and accountability: Covid 19**

Thank you very much for your letter following my appearance at the Committee meeting on 24th November. I am happy to provide some answers to your questions.

**1. Can you give a view on whether the rate of infection can be inferred from the numbers of people testing positive, given these limitations?**

Based solely on the testing positive figures, I would say no – the rate of infection cannot be inferred. There is information in the testing positive figures, but this information needs to be calibrated using, at a minimum, concurrent information on the total number of tests and, better still, the results from the designed surveys.

As explained in my witness statement, the testing positive figures are influenced by the number of tests performed each day. The number of tests itself is influenced not only by the current force of infection but also by the availability of tests, existence of a local information campaign or a local mass testing exercise, possibly socio-demographic factors, weekly patterns etc. So, at minimum, one needs to interpret the swab positive figures ***jointly*** with the corresponding testing information: i.e., to know out of how many tests these figures come from, and metadata which would capture the existence of specific local testing circumstances.

What is important is to monitor the positivity rate, i.e., to compute the number of positive swabs amongst the swabs performed on a specific date. This information on positivity is quoted as an indicator in the criteria that have been listed by HM Government for allocation of tiers. It is regrettable that such information is not yet openly accessible for analysis (as far as I know), as that would allow a clearer understanding of why areas are allocated to specific tiers. Ideally it would be computed at a fairly low spatial resolution such as Lower Tier Local Authority – and it would be helpful to know what resolution is being used.

Note that it takes a few days to have the results from the tests, so the number of positive swabs reported on one day is different from the number positive swabs *occurring* on that day, which is the more interpretable quantity. This is a well-known issue for deaths because of the long delay with processing death certificates, but the same issue is also present for the test results. Hence stable estimates will need to allow for a few days lag. Tracking very recent time evolution cannot be done on the raw numbers and requires additional nowcasting techniques.

On a different note, it is important to remember that people coming for tests *do not represent a random sample of the population*, so the swab positive figures need to be compared with the prevalence rates that can be estimated from the REACT and ONS representative surveys to understand and quantify the types of biases present in the pillar 2 data. As far as I know this calibration has not been done in details, hence the rate of infection cannot be reliably inferred from the testing figures and the two surveys remain essential for monitoring the rate of infection.

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

12 Errol Street
London EC1Y 8LX

020 7638 8998
rss@rss.org.uk
www.rss.org.uk

In conclusion, monitoring the pillar 2 positive test figures and the positivity rate in a local area can give a partial indicator of the local trend but in the absence of a detailed calibration study with respect to the designed random surveys, these figures need to be interpreted with caution.

**2. Can you tell us what research tells us about the effectiveness of NPIs such as wearing masks and social distancing? Which measures might work best, and which might be less effective? Can you also give a view on whether the Government has data available to make a judgement on NPIs, is making the best use of that data, and is collecting data to fill gaps?**

***International studies***

There are a large number of studies that have addressed the evaluation of the role of Non-Pharmaceutical Interventions (NPIs) in a variety of ways, in particular *using information from different countries*, where *different sets of NPIs* have been implemented *at different times in the epidemic curve*, thus providing informative variability that can be usefully statistically analysed. Assessing the quality of such studies and summarising the evidence is best done by expert groups such as SAGE as there is a large body of literature using a variety of methods. **In my answers I am only highlighting a few recent studies without attempting to be comprehensive, please take my comments in this spirit.**

A recent comprehensive study published in the *Lancet* is of particular interest. In this study, time varying R is linked to the introduction and lifting of NPIs using data from 131 countries. Overall, the study found *time delayed influence of a number of NPIs on decreasing trends in R*, of varying degree following the type of NPIs.

Specifically, the researchers linked data on daily country-level estimates of R from the London School of Hygiene & Tropical Medicine (LSHTM) with data on country-specific policies on NPIs from the Oxford COVID-19 Government Response Tracker, available between Jan 1 and July 20, 2020. The researchers conclude that individual NPIs (including school closure, workplace closure, public events bans, bans on gatherings of more than ten people, requirements to stay at home, and limitations on movement within a country) are associated with reduced transmission of the virus, but the effect of introducing and lifting these NPIs is delayed by 1–3 weeks. The research also suggests that the effect is delayed more when lifting NPIs than when introducing them.

A number of earlier studies had previously used similar methods to investigate the effect of imposing NPIs during the first period of lockdown. The Lancet study linked to above, considers a longer time interval and also investigates the effect of lifting NPIs.

Comment on whether the Government has data available to make a judgment on NPIs: There is now a solid body of literature including some of the recent studies that I have highlighted. Analysing the UK data with similar methods to some of these papers may be possible but with much smaller geographical areas, there would be considerable additional noise and uncontrolled sources of variability based on local factors which would make quantifying the effect of local NPIs more difficult. For example., the testing data that would need to be used as a basis for tracking the evolution of local R rates over time is highly variable. Hence this would require time and effort and a considerable input of statistical expertise to be reliably done, as well as the availability of good data pipelines.

***Social distancing***

Evaluation of measures related to social distancing is included in all the general studies of NPIs.

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

12 Errol Street
London EC1Y 8LX

020 7638 8998
rss@rss.org.uk
www.rss.org.uk

Of note is the *BMJ* study by Islam et al which investigated physical distancing interventions (closures of schools, workplaces, and public transport, restrictions on mass gatherings and public events, and restrictions on movement (lockdowns)) between 1 January and 30 May 2020 in 149 countries.

They found that on average, implementation of any physical distancing intervention was associated with an overall reduction in covid-19 incidence of 13% (IRR 0.87, 95% confidence interval 0.85 to 0.89; n=149 countries).

Another recent study has been published in *Nature Human Behaviour* by Haug et al which ranks the effectiveness of worldwide Covid-19 government interventions. Social distancing, in particular the cancellation of small gatherings, comes out consistently as the most effective (Fig1).

***Masks***

Mask wearing has primarily been discussed as a means of controlling transmission. There are a number of systematic reviews and meta-analysis with compelling evidence, e.g., quoted in a recent editorial by Frieden and Goldwasser, that masks are effective for reducing transmission of the virus..

The Royal Society have also published a review of evidence on facemasks. Basically, using a face mask reduces transmission through droplets – one of the major routes (but not the only one) of transmission – and this is particularly relevant for Covid-19 since there is transmission of SARS-CoV-2 by asymptomatic and pre-symptomatic individuals, which are not targeted through measures such as isolation etc.

Based on ecological evidence (e.g., Hong Kong, where they had R=2 despite universal mask wearing), we would expect a maximum of 30-40% reduction in overall transmission if everyone wore a mask in public, which is very significant in terms of public health.

The evidence that mask wearing would protect the individual wearing it – i.e. that it contributes to reducing *infection risk* – has been the subject of discussion. It is of course subject to duration of wear, context etc. One would not expect the reduction only in infection risk to be more than 20%. The recent Danish study by Bungaard et al was designed to test for 50%+ reduction in infection risk, so a smaller reduction (e.g., 10-20%) wouldn't show up as significant – even though, in terms of public health, it would be a very useful effect to have at population level. As discussed in the editorial by Frieden and Goldwasser, the Bungaard paper had a number of methodological issues. There was poor self-reported mask wearing compliance rate of 46%, and the serological test used was poorly specific – both facts would bias the results towards the null. So, it is a *pretty inconclusive study* and it should not be used to say that mask-wearing is ineffective for preventing the spread of the virus as some UK commentators have implied.

Finally, the insightful paper in the *BMJ* by Marteau and colleagues, also lists a number of reviews. This paper presents a convincing argument against the claim that mask-wearing would increase risk compensation (i.e., cause people to take other risks because they feel safer by wearing a mask) .

**3. You told us that spreading events (i.e. gatherings of people) played a key part in transmission of infection. Can you briefly outline what this information tells us about the role that gatherings play, and what it tells us about the transmission risk in different settings?**

There are a number of studies (see below for some) that clearly point towards the importance of what is commonly referred as "**superspreading events (SSE)**", i.e.. patterns of transmission where a small fraction of infected people create a larger number of infections. Currently the estimate ranges from 5 to 20% of cases seeding up to 80% of infections.

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

12 Errol Street
London EC1Y 8LX

020 7638 8998
rss@rss.org.uk
www.rss.org.uk

Althouse et al's article on transmission dynamics from May 2020 is an early discussion of the super-spreading concept and its implications.

This paper argues that: "The basic reproduction number, which has been widely used and misused to characterize the transmissibility of the virus, hides the fact that transmission is stochastic, is dominated by a small number of individuals, and is driven by super-spreading events (SSEs). The distinct transmission features, such as high stochasticity under low prevalence, and the central role played by SSEs on transmission dynamics, should not be overlooked. Many explosive SSEs have occurred in indoor settings stoking the pandemic and shaping its spread, such as long-term care facilities, prisons, meat-packing plants, fish factories, cruise ships, family gatherings, parties and night clubs. These SSEs demonstrate the urgent need to understand routes of transmission, while posing an opportunity that outbreak can be effectively contained with targeted interventions to eliminate SSEs. Here, we describe the potential types of SSEs, how they influence transmission, and give recommendations for control of SARS-CoV-2."

Endo et al (July 2020) explain the statistical basis of overdispersion in the number of secondary infections and quantify it using outbreak sizes in different affected countries. They summarise their findings as follows: "While sustained transmission chains of human-to-human transmission suggest high basic reproduction number $R_0$, variation in the number of secondary transmissions (often characterised by so-called superspreading events) may be large as some countries have observed fewer local transmissions than others…. Our model suggested a high degree of individual-level variation in the transmission of COVID-19. Within the current consensus range of $R_0$ (2-3), the overdispersion parameter $k$ of a negative-binomial distribution was estimated to be around 0.1 (median estimate 0.1; 95% CrI: 0.05-0.2 for R0 = 2.5), suggesting that 80% of secondary transmissions may have been caused by a small fraction of infectious individuals (~10%). **Conclusions:** Our finding of a highly-overdispersed offspring distribution highlights a potential benefit to focusing intervention efforts on superspreading. As most infected individuals do not contribute to the expansion of an epidemic, the effective reproduction number could be drastically reduced by preventing relatively rare superspreading events."

Risk factors for the occurrence of SSE are a combination of biological (e.g., high viral load) and behavioural (e.g., many contacts). It is likely that pre-symptomatic transmission plays an important role in SSEs.

**Social contexts**: Some social or work contexts are favourable to these superspreading events as they tend to involve having a large number of people in close proximity.

Evidence of the origin of SSE events has mostly be derived by carrying out detailed ***back tracing of clusters.*** Below I quote a few insightful examples, there are many others.

The recently published study by Adam et al in *Nature Medicine*, which points to large clusters in Hong Kong being linked to social venues such as bars, restaurants, weddings and religious sites, is a good example of the evidence provided by detailed cluster analysis. This study investigated contact tracing data to characterize clusters (≥2 cases) of SARS-CoV-2 infections associated with 1,038 confirmed SARS-CoV-2 infections in Hong Kong between January 23 and April 28, 2020. An investigation of 137 different recognized clusters (median cluster size = 2) found that 7 probable SSEs accounted for 58% of all clustered cases.

- The largest cluster of 106 cases was associated with four bars in Hong Kong.
- An estimated 19% (95% CI: 15% – 24%) of cases caused 80% of all local transmission.
- Transmission was most frequently observed within family households (92/169, 54.4%), followed by social (56/169, 33.1%) and work (20/169, 11.8%) settings

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

12 Errol Street
London EC1Y 8LX

020 7638 8998
rss@rss.org.uk
www.rss.org.uk

- Conclusion: There is substantial potential for SARS-CoV-2 superspreading in settings where large numbers of people gather such as bars, weddings, and religious events. Interventions targeting social settings may be key in reducing the risk of SSEs and SARS-CoV-2 transmission.

Another specific example is detailed in Mahale et al's article 'Multiple COVID-19 Outbreaks Linked to a Wedding Reception in Rural Maine'. This details how a wedding reception with 55 people led to Covid-19 outbreaks in the local community, as well as at a long-term care facility and a correctional facility in other counties. Overall, 177 Covid-19 cases were linked to the event, including seven hospitalisations and seven deaths (four in people who were hospitalised). The investigation revealed noncompliance with CDC's recommended mitigation measures.

Besides analysis of clusters, other types of studies have been carried out to study the influence of social contexts, such as restaurants. Because of its case control design, and its size, a recent study published by CDC (Fisher et al, September 2020) is of particular interest. Here, a case-control investigation of symptomatic outpatients from 11 US health care facilities found that both close contact with people known to have the virus and visiting places like restaurants and bars were associated with Covid-19 positivity. Adults with positive SARS-CoV-2 test results were approximately twice as likely to have reported dining at a restaurant than were those with negative SARS-CoV-2 test results. In particular, cases without close contacts to known confirmed Covid cases, were more likely to have been to a restaurant (OR 2.8) or a bar/coffee shop (OR 3.9).

There is also a US modelling study using mobile phone mobility data to look at fine scale social interactions that points towards the influence of visits to points of interest such as restaurants and religious establishments on infection rates. **()**

In the UK, there has been an investigation of the effect of the "eat out to help out" scheme in England, 'Subsidising the spread of Covid-19', by Fetzer, was recently published as a preprint. This concludes that between 8-17% of new infections may have been caused by 'eat out to help out'. Areas with higher take-up saw both a notable increase in new Covid-19 infection clusters within a week of the scheme starting and, again, a deceleration in infections within two weeks of the program ending**.**

Finally it is worth highlighting the database of international media reports on SSEs from LSHTM.

I hope you find these answers helpful. Please do let me or staff at the RSS know if there is anything else we can do to help.

Yours sincerely,

Professor Sylvia Richardson CBE
President-Elect of the Royal Statistical Society and Co-Chair of the RSS Covid-19 Task Force