

Written evidence from the Royal Statistical Society (DTA 42)

Public Administration and Constitutional Affairs Committee Data Transparency and Accountability: Covid 19

Introduction

This is the submission from the Royal Statistical Society (RSS) to the Public Administration and Constitutional Affairs Committee's inquiry into 'Data transparency and accountability: Covid-19'.

We are focusing on questions 1, 4 and 6, since they link most closely to the Society's longstanding concerns around the importance of transparency and public understanding of statistics.

In our response to these questions we make the following recommendations:

1. That a formal government review, analogous to the Bean review of economic statistics, is conducted into England's health data to ensure that a well-functioning system is established at the earliest opportunity.
2. Whenever data is referred to in support of a decision, the government should publish all relevant data alongside the decision and clearly signpost to it.
3. A considerable strengthening of Office of Statistics Regulation: increasing its resources to a level at which it is able to proactively identify issues with the production of statistics and reviewing its powers to see if there is a case for strengthening them.
4. The UK Government should publish a clear framework for how data will be used to make decisions regarding local, regional or national lockdowns in England and should publish the relevant data when making decisions.
5. If regular briefings are required again for England, a mechanism should be introduced to ensure independent and non-political communication of data – such as a weekly briefing to journalists by the national statistician, chief medical officer or chief scientific officer.

We believe that, if implemented, these recommendations would together substantially improve both public confidence in and understanding of the UK Government's decision-making process.

1. Did Government have good enough data to make decisions in response to Coronavirus, and how quickly were Government able to gather new data?

1.1. To answer this question, it is helpful to set out the RSS's view on what it would mean for the Government to have "good enough" data.

1.2. At the start of the pandemic in early February 2020 the UK Government had epidemiological evidence coming from China of sustained person to person transmission of SARS-CoV-2. Evidence on main transmission routes – on characteristics of the transmission dynamic and on case fatality ratios, informed by data mostly coming from China – were [presented to Government by its Scientific Advisory Group in Emergencies](#) (SAGE). SPI-M (Scientific Pandemic Influenza-Modelling), subgroup of SAGE quickly adapted its infectious disease modelling to address SARS-CoV-2. By using international data on incubation period, infectiousness and fatality-rates, projection scenarios for Covid-19 hospitalizations and fatalities were available in early March 2020.

1.3. As soon as it was known that some of the characteristics of the transmission dynamics were markedly different from previous viruses, such as the probability of pre-symptomatic and asymptomatic transmission, it was clear that large-scale active surveillance and data collection programmes were necessary to inform about the spread of the virus. The countries that moved quickest in setting up such programmes were those that had experienced the SARS-CoV 2003 epidemic (such as China, Hong Kong, Singapore, Vietnam, South Korea).

1.4. So, in the early phase of an epidemic, to have good enough data means having data-informed infectious disease models and using these to make decisions. However, as the situation develops, priority has to be given to improving data by setting up a comprehensive surveillance system, including designed studies for active surveillance.

1.5. The RSS's view is that good enough data was available early in the pandemic, in the sense that the UK Government had access to data-informed models. However, it is less clear that the scale of surveillance required to provide good quality data to track the unfolding epidemic in the UK was achieved as quickly as it could have been. In order to assess the speed at which the Government collected new data, it is important to consider structural challenges faced by statisticians, while highlighting both key achievements and improvements that could have been made.

Structural challenges

1.6. There was a series of structural challenges facing government statisticians at the start of the pandemic – each of which made the process of generating, analysing and making use of good data difficult:

1.6.1. Because health and social care data in the UK is devolved and a variety of organisations produce data – each of the four nations of the UK has data collection split between its government, NHS, civil registration agency and public health body – coherent UK-wide data requires a level of collaboration and communication which is difficult at the best of times, and harder still in the pressure of a pandemic. In addition, the different demands of the devolved administrations mean that data and analytics tend to be focussed on the needs of the devolved nation rather than the UK as a whole

1.6.2. In England, NHS England and its data collection function have been fragmented into multiple agencies. The Office for Statistics Regulation (OSR) [published a systemic review of health and social care statistics in England in 2015](#) and

concluded that “there was no single individual or organisation with clear leadership responsibility and this had led to problems with the coherence and accessibility of these statistics”. Progress has been made, but effort is still needed to maintain that improvement – and this is difficult during a pandemic.

1.6.3. Because of this fragmentation in England, statisticians and data analysts are spread throughout the health system and there is a shortage of statisticians centrally in the Department for Health and Social Care (DHSC), where they were needed to pull together data from this disparate array of sources. The situation has been better in Scotland, where the vast majority of health statistics relating to COVID-19 are produced by Public Health Scotland (formerly ISD (Information Services Division) Scotland and Health Protection Scotland).

1.6.4. The daily press briefings generated a pressure for real-time data on hospital admissions and deaths. In England, at the outset, there was a daily-reported number of confirmed Covid-19 hospitalised deaths, although using report-date meant inevitable delays and what was needed to inform decisions was an understanding of the number of deaths that *occurred* – rather than that were reported – on a given day *in all settings*. This required modelling and accounting for the reporting delays.

1.7. Faced with these formidable challenges, the Government Statistical Service (GSS) and UK Statistics Authority (UKSA) have devoted a lot of energy to proactively identify and tackle them: effectively adapting existing systems, introducing new systems rapidly to improve data infrastructure (for example linkage of death registrations with census and Hospital Episode Statistics data to explore the role of ethnicity and health in Covid outcomes) and producing a series of informative and detailed reports.

1.8. In Scotland, the merger of NHS and public health data collection in the new Public Health Scotland has been hugely beneficial for the sharing of resources and skills relating to data management, analysis and interpretation

1.9. While in this respect the GSS and UKSA deserve praise for their responses to Covid-19, there is a question about their readiness. In 2016, the [Bean Review of economic statistics](#) recommended that the Office for National Statistics (ONS) should “move away from focusing largely on the production of statistics and become more of a service provider, helping users answer their questions about the economy” (p.10). While this review focused on economic statistics, many of its findings could equally have been applied to health statistics. We believe that a similar review into health statistics, building on the OSR’s review of English health data, would be beneficial.

Recommendation 1: The RSS recommends that a formal government review, analogous to the Bean review of economic statistics, is conducted into England’s health data to ensure that a well-functioning system is established at the earliest opportunity.

1.10. This review should involve health, data science and statistical experts from outside the current structures. We would suggest that it should be conducted as a “no fault” review

that focuses on processes, streamlining, efficacy and outcomes. The review should look at the entire data pipeline and its underpinnings, including staffing and education.

Achievements

1.11. As indicated above, the performance of UKSA meant that the UK Government was able to gather some new data, covering England, reasonably quickly. The speed with which the ONS was able to increase its production of health statistics is commendable, and the manner in which it has engaged with departmental producers of statistics has clearly materially improved both the collection and presentation of data.

1.11.1. The ONS Covid-19 Infection Survey and Imperial's REACT programme have been particularly important sources of data to inform the government's response. Both the ONS Infection Survey and REACT-1 were designed in remarkably quick time with fieldwork begun in May and June 2020. The ONS Infection Survey has expanded from its pilot phase, but expansion required a major change in sampling frame with consequent decrease in volunteer-rate. REACT-1 uses a different sampling frame, and aims to recruit over 100,000 participants per round so that its regional estimates are suitably precise. Its [timely latest report](#) is based on 86,000 individuals in round 6 and >770,000 in rounds 1 to 5. Both also measure infection-incidence, ONS Infection Survey directly, REACT-1 indirectly.

1.11.2. Both ONS Infection Survey and REACT-2, using different antibody tests, also monitor antibody prevalence. REACT-2, in particular, has used its prevalence data to infer the epidemic curve.

1.11.3. Antibody prevalence measured in blood donors up to June 2020 have informed nowcasting and forecasting during UK's first wave of SARS-CoV-2 infections.¹

1.11.4. UK's Medicines and Healthcare products Regulatory Agency (MHRA) does not, in general, licence in-vitro diagnostic kits for infectious diseases but MHRA took a lead, with helpful input from RSS, by [publishing Target Product Profiles](#) to guide manufacturers to have a robust basis for self-certification of SARS-CoV-2 antigen or antibody tests.

1.12. Although both the ONS and REACT programmes are successful on their own terms a single, larger programme dedicated to real time surveillance would have enabled the government to gather more data more quickly. The RSS view is that this should have been a higher priority. We welcome the recent establishment of six National Core Studies but would like to stress that attention should be paid to add purposely designed data collection within these studies in order to help statistical synthesis of key quantities of interest across core studies. Regrouping different studies under one umbrella is only the first step.

¹ See Birrell et al (2020), [Real-time Nowcasting and Forecasting of COVID-19 Dynamics in England: the first wave?](#)

Areas for improvement

1.13. There are specific areas where improvements have been necessary, and in some cases there is still some work to be done:

- 1.13.1. The initial collection and presentation of Test and Trace (T&T) data was not of the necessary standard. The statistician-team at DHSC who were charged with producing official statistics on T&T operations had no input to the design or inter-operability of the data-collection systems from which data needed to be extracted to monitor operational matters. Difficulties were even greater when it came to answering important infection control questions that statisticians, public health specialists and infectious disease modellers would expect to be answered. Accordingly, on 23 July 2020, the RSS Covid-19 Taskforce issued a [statement on using the Test and Trace data to increase knowledge on transmission](#), in particular through linkage (which the RSS has called for greater use of) and on how the T&T system could be utilised to learn about adherence to and effectiveness of self-isolation as it is currently practised.
- 1.13.2. However, there have been continued improvements in data presentation – the RSS has held regular meetings with DHSC staff who are clearly committed to improving the reporting and DHSC staff have also engaged positively with UKSA. The Public Health England (PHE) dashboard has steadily improved in quality and depth, and now forefronts positive tests and deaths by actual day of occurrence rather than reporting, while making all the data available to programmers through an API.
- 1.13.3. While improvements have been made, some poor data presentation persists: the daily “tested positive” figure, which is not helpful without knowing who has been tested and why, and diagrams of test-results against **date** which do not specify which date the axis refers to, a clear case of poor practice; there is still no report of the positivity rate, an important indicator, nationally, regionally and locally.
- 1.13.4. In general, when data-driven estimates and model-based predictions have been presented, there has been a lack of proper uncertainty quantification – which makes the estimates and predictions impossible to fully interpret.
- 1.13.5. The initial definition used by PHE of a death due to Covid-19 – which counted any death after a positive test for the virus as a death due to the virus, regardless of the time that had passed since the positive test – should very quickly have been seen as inadequate. This has now been corrected, but it was an oversight that could have been addressed earlier in the process. It is also noteworthy that there is no standard internationally accepted definition of a death due to Covid-19: as the ONS is a leader in harmonisation of economic statistics, there is an opportunity here for the UK to lead on harmonisation of Covid-19 and public health statistics.
- 1.13.6. Comprehensive indicators of the UK preparedness for the winter in terms of healthcare capacities and procurement, including data on stocks of PPE, have not been made widely available despite [early calls from professional bodies such as](#)

[Academy of Medical Sciences](#) that winter preparedness needed to be comprehensively tackled.

4. Were key decisions (such as the “lock downs”) underpinned by good data and was data-led decision-making timely, clear and transparently presented to the public?

4.1. When managing a public health emergency in real-time, the decision process is continually evolving and what is important is to strive to make the best decision given the available evidence, while highlighting areas of uncertainties – decision processes in public health emergencies are different from those used to approve therapies. This is precisely why it is of paramount importance to be transparent about any assumptions that are made, what evidence from the literature is used, what are available data that support/dispute the decision and which data are lacking: where there is uncertainty it should be acknowledged. This will help the scientific and public health community to identify priorities for collecting data as well as build public trust.

4.2. Transparency about evidence is vital because:

4.2.1. It makes it easier for people to meaningfully assess government’s decisions – as well as being important as a democratic principle, this process also leads to better decision-making, by opening the process to constructive criticism.

4.2.2. Transparency around data can help the public understand why decisions have been made and make it more likely that people will respect and abide by those decisions. Part of the reason that the initial lockdown was widely adhered to was that the presentation of data helped people to clearly see the severity of the situation.

4.3. The press briefing announcing the second lockdown on October 31st provides a good example of both good and poor practice. The slides shown by the Chief Medical Officer and the Chief Scientific Advisor were made immediately available, and most of the sources for the data were given. However, a slide showing projections of deaths from different models was not only based on reports by SPI-M that were not available on October 31, but also used working analysis and scenarios dated from the week beginning October 9th, while more recent data driven forecasts were available to better describe the upward trend supporting the decision of a second lockdown. This was repeated in the [briefing pack given to MPs](#) ahead of the vote on new regulations on 4 November. In order not to undermine public confidence, we recommend that the latest forecasts with uncertainty quantification are used whenever available.

4.4. Lack of transparency has been a consistent problem in the government’s management of Covid-19. For example, in August, the Prime Minister announced that home and garden visits would be made illegal in parts of northern England, claiming that the data showed this to be the main source of transmission in those areas – however this data was not published at the time of the decision and so it was impossible to see whether this was the case. The [RSS commented publicly on this at the time](#).

4.5. There has also been a lack of transparency around some aspects of the tiers-system for local areas. First, any data that has informed decisions about what activities are permitted and which types of business can operate at each tier has not been made available ahead of the decision being taken. Second, there is a lack of clarity around what determines the risk level in any given region: what is the set of indicators and the process undertaken before a decision is made to move an area between tiers?

Recommendation 2: The RSS recommends that whenever data is referred to in support of a decision, the government should publish all relevant data alongside the decision and clearly signpost to it.

4.6. There is also a question around how scientific advice is handled. The Government has been slow to publish minutes of meetings and the advice that they have received: we believe that during a public health emergency it is especially important that the Government is open with the public about the advice they are receiving and, if it is not followed, they should give a clear explanation. This would also prevent advice leaking, and make it easier for the government to control public health messaging.

4.7. The OSR has, throughout, performed a valuable role in pressing for greater transparency of the data used to inform decisions. This has led to some improvements, but there is still room for greater transparency and timeliness. It is our view, in line with the Public Administration and Constitutional Affairs Committee's recommendations in [last year's review](#), that that OSR needs a considerable increase in resources. However, we also believe that OSR's legal and enforcement powers need to be reviewed and strengthened.

Recommendation 3: The RSS recommends a considerable strengthening of Office of Statistics Regulation: increasing its resources to a level at which it is able to proactively identify issues with the production of statistics and reviewing its powers to see if there is a case for strengthening them.

4.8. It is also worth drawing attention to some of the good practice in Scotland. The Scottish Government still holds daily briefings – led by the First Minister – and any data referred to at these briefings are published alongside the briefing. The Scottish Government has [also published a framework for decision making in the context of Covid-19](#) –regularly reviewing the situation and publishing the data informing their decisions.

Recommendation 4: The UK Government, like the Scottish Government, should publish a clear framework for how data will be used to make decisions regarding local, regional or national lockdowns in England and should publish the relevant data when making decisions.

6. Is the public able to comprehend the data published during the pandemic. Is there sufficient understanding among journalists and parliamentarians to enable them to present and interpret data accurately, and ask informed questions of Government? What could be done to improve understanding and who could take responsibility for this?

6.1. Journalism, supported by the Science Media Centre, has played a positive role during the pandemic. In our experience journalists have made serious and timely efforts to understand and accurately present data. This has been all the more necessary because the UK

Government's communications operation has not always acted in a way that prioritised public understanding of the data.

6.2. As well as the Science Media Centre, Sense about Science has also had a positive impact. Both organisations through informal departmental channels and their network of scientific commentators, have helped journalists to present data accurately.

6.3. A number of RSS Fellows have engaged regularly and successfully with journalists – helping them to understand the data well enough to offer accurate explanations to the public. Specialist journalists – health or science and technology correspondents – have been generally very good at reporting data accurately, including graphically – and the BBC, for example, has benefitted through having a Head of Statistics.

6.4. The situation is somewhat less positive when considering journalists who do not routinely report on health or science. Political journalists, who tended to be the ones invited to ask questions at the daily briefings, did not often ask the sort of incisive question about data that could have been helpful (though the format of the briefings did not really lend itself to that in any case), but that may have been because there were other – political – issues that they were more interested in rather than reflecting a lack of understanding.

6.5. At times it has seemed that the presentation of statistics has been impacted by political considerations. There are two particularly stark examples. First, at the start of the crisis the UK Government set a target to carry out 100,000 Covid-19 tests per day by the end of April – and the Government claimed to have met this target through posting over 100,000 tests on one day. The RSS issued a [statement on this in April](#) and it also resulted in the chair of UKSA writing to the Secretary of State for Health and Social Care to emphasise the importance with adhering to the Code of Practice for Statistics. While some improvements have been made in the presentation of the figures around testing, there remains a focus on achieving large numbers without a clear statement of the government's objective: this is a point the RSS has made in September in a [letter to the Times](#).

6.6. Second, at the daily briefings – held at the height of the crisis – daily death figures were announced. The co-chair of the RSS's Covid-19 Task Force, Sir David Spiegelhalter, has described these briefings as “number theatre”: large numbers were presented that were both inaccurate and overly-precise. For example, the number of deaths (for the most part, deaths reported on a day in a hospital setting) presented was almost always an under-estimate and the precision of the number presented gave a false impression of certainty.

6.7. UKSA had staff embedded who helped to improve how the information was presented. However, it needed more than this: the people presenting the data needed to be comfortable in explaining clearly what the data did and did not say, and where there were uncertainties.

Recommendation 5: The RSS recommends that, if regular briefings are required again for England, a mechanism is introduced to ensure independent and non-political communication of data – such as a weekly briefing to journalists by the national statistician, chief medical officer or chief scientific officer.

November 2020