
Chair of the UK Statistics Authority, Sir David Norgrove

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

9 February 2021

Dear William,

Thank you for your letter of 2 February, in which you asked about the Government's use of data, in relation to your inquiry *Data transparency and accountability: COVID-19*.

We discuss your specific questions in an Annex to this letter. Here I would like to reflect more generally about the use of data and statistics in the past year.

Overall I believe our statistical system has responded well to the stress and pressures of the pandemic. Ian Diamond's separate letter to you describes an immense range of work that has been done to understand the pandemic itself, which has been fundamental to government decision making and public understanding. Alongside the work on the pandemic the Office for National Statistics (ONS) and statisticians across government have continued to produce remarkable data and analysis across the economy and society, work that is high quality and innovative. Preparations and contingency plans for the census in England and Wales are encouraging.

The legislative framework for our statistics as set out in the Statistics and Registration Service Act 2007 together with the Digital Economy Act 2017 has also I think met the sternest test it has yet seen. The new data and statistics required by the pandemic have for the most part been compiled and published in accordance with the Code of Practice on Statistics and statisticians have generally been able to access the new sources of data they need.

I pay warm tribute to all involved in this work, at a time of anxiety for them and their families, with all the disruption caused by the need to work from home, alongside the increased difficulty of their professional lives, with many surveys and other sources of data having to be changed or abandoned.

Within this generally positive picture not all has gone well, and there are lessons to be learned.

It has too often been a struggle to develop a coherent picture of the pandemic even within England as a single country. DHSC currently plays a limited role in health statistics. Its resource has been strengthened following an ONS review undertaken at their request. But the disparate bodies involved in the provision of health are in terms of statistical output too often inchoate, to the extent for example that both the NHS and Public Health England produce statistics on vaccinations that are published separately.

This is an issue that has been highlighted by the Office for Statistics Regulation (OSR) in the past. It goes well beyond the concerns raised by the pandemic. We currently have no

coherent statistical picture of health in England or of the provision of health services and social care.

There are similar issues in relation to the health data for the four nations. The adoption of different definitions complicates comparisons and makes it hard to draw the valuable lessons we could all learn from different ways of doing things.

I strongly support the proposal by the Royal Statistical Society for a thorough external review.

More immediately it is hard to understand why the different nations have chosen to publish vaccination data in the different ways and detail they have chosen. OSR is pursuing this.

Some people may be surprised by my mostly positive assessment of the handling of statistics and data over the past year. Their more negative view is likely to have been influenced by a number of – too many – particular examples of poor practice.

- The presentation of data at No 10 press briefings has improved, helped by the later involvement of ONS staff, but early presentations were not always clear or well founded, and more recently a rushed presentation has undermined confidence.
- Ministers have sometimes quoted unpublished management information, and continue to do so, against the requirements of the Code of Practice. Such use of unpublished data leads of course to accusations of cooking the books or cherry picking the data. It should not require my involvement or that of OSR to secure publication.
- Perhaps most important is the damage to trust from the mishandling of testing data. The target of 100,000 tests per day was achieved by adding tests sent out to tests completed. As predicted there was huge double counting, to the extent of some 1.3 million tests that were eventually removed from the figures, in August. The controversy over testing data seems likely to continue to undermine the credibility of statistics and the use that politicians make of them.

The Annex describes a range of current issues in relation to the pandemic, including testing and vaccinations, as well as replying more directly to your letter.

There are perhaps two areas the Committee might like to consider in terms of future change.

The first is the central role of the Authority together with the National Statistician and OSR. The UK has a decentralised system of statistics where individual departments are responsible for their statistics and departmental statisticians report within their departments. This has strengths we should not lose. It ties statistics and statisticians closely into the policy making of their departments and any change should not weaken that tie. But the complexity of data and statistics in the current crisis has shown the need in these circumstances for a firmer central controlling mind. The National Statistician and the ONS have taken this role to a large extent, through expertise, position and personality rather than formal agreement. OSR has also taken a more expansive role. For the future there may be a place for more formal arrangements.

Secondly, it is clear that political pressures have led to some of the weaknesses in the handling of Covid statistics. It is to the credit of our politicians that they have created an organisation like the Authority that is permitted to criticise them, and in general politicians respond appropriately to our criticisms. But it might help if more issues were headed off

before they arose. The Ministerial Code for example only asks Ministers to be 'mindful' of the Code of Practice. The requirement could be stronger.

The Authority in 2020 published a new five year strategy¹, during the pandemic. It remains valid and we are pursuing it at pace. The ONS is leading the development of an Integrated Data Platform for Government as well as developing new and better statistics to help the country understand the economy and society, from trade to happiness and from crime to migration. Statistics to help the recovery are a particular focus. OSR is developing its work on statistical models – its review of exams algorithms will be published shortly – as well as on automation of statistics, data linkage, National Statistics designation, granularity and statistical literacy.

I look forward to keeping you and the Committee in touch with our progress.

Kind regards

A handwritten signature in black ink that reads "David". The letters are cursive and slightly slanted to the right.

Sir David Norgrove

¹ <https://uksa.statisticsauthority.gov.uk/statistics-for-the-public-good/>

OFFICE FOR STATISTICS REGULATION ANNEX

1. Has the Government made enough progress on data since the start of the pandemic, and what gaps still remain?

Summary

Since the start of the pandemic Governments across the UK have maintained a flow of data which has been quite remarkable. New data collections have been established swiftly, existing collections have been amended or added to, and data sources have been linked together in new ways to provide additional insight. We have seen good examples from across the UK, including data on the virus itself and on the wider impacts on the pandemic.

However, in some areas Governments have been slow both to publish data and to ensure its fitness for purpose. For example, more consideration should have been given to the data available as testing and tracing was being set up. While UK governments have started publishing vaccinations data more promptly, and are continuing to develop the statistics on an ongoing basis, there remains much room for improvement in terms of the amount of information that is published and the breakdowns within the data. There are also gaps remaining in the data available to support longer term understanding of the implications of the pandemic.

It is clear that there is intensive work taking place to provide more comprehensive vaccinations data, both by each Government, and through cross-UK collaboration.

New data developed

There are many examples of outputs which have been developed quickly to provide new insights to help understand the pandemic and its implications. Some specific examples include:

- The coronavirus (COVID-19) infection survey, carried out by the Office for National Statistics (ONS) in conjunction with partners. This is the largest and only representative survey in the world that follows participants longitudinally over a period of up to 16 months.
- Statistics on the Coronavirus Job Retention Scheme (CJRS), the Self-Employment Income Support Scheme (SEISS) published by HM Revenue and Customs (HMRC), and the ONS Business Impact of Coronavirus Survey (BICS).
- The Ministry of Justice (MoJ) and Her Majesty's Prison and Probation Service (HMPPS) have published official statistics providing data on COVID-19 in HM Prison and Probation Service in England and Wales.

We have also seen outputs which attempt to draw together data to make it more easily accessible. The most well-known of these is the coronavirus dashboard² which is widely used and constantly evolving. While the dashboard focuses on data related to COVID-19 infections, hospitalisations and deaths, the ONS has pulled together a broader range of data from across the UK government and devolved administrations to highlight the effects of the pandemic on the economy and society: *Coronavirus (COVID-19): 2020 in charts*³. This output covers a broad range of data including transport use, schools attendance and anxiety levels.

² <https://coronavirus.data.gov.uk/>

³ <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid192020incharts/2020-12-18>

Coronavirus (COVID-19) data

We have seen improvements to the data since the start of the pandemic, particularly around cases and deaths, with clearer information on the different sources of data available and the value of each of these sources.

However, we continue to have concerns, including seeing examples of data being referenced publicly before they have been released in an orderly manner, which we have highlighted to your Committee. In some areas Governments have been slow both to publish data and to ensure its fitness for purpose. For example, the UK coronavirus dashboard is now a rich source of information with plans for the inclusion of further data. However, it took several months to become the comprehensive source it is now.

Our concerns around COVID-19 health data cover three broad areas: *Testing and tracing, hospitalisations, and vaccinations.*

Our concerns with data on *testing* have been made public on a number of occasions⁴. While there have been significant improvements to the England Test and Trace data it is still disappointing that there is no clear view of the end to end effectiveness of the test and trace programme.

In December we published a statement on data on *hospital capacity and occupancy*⁵, noting the biggest limitation to the existing data is the inability to make a distinction between whether someone is in hospital because of COVID-19 infection, or whether they are in hospital for some other reason but also have a COVID-19 infection. These data should become available in time but the delay limits understanding at a critical time.

On *vaccinations data*, UK governments have been quick to start publishing data and have learnt some of the lessons from test and trace (see section 4). However, there remains room for much improvement in terms of the amount of information that is published and the breakdowns within the data. We would like to see more granular breakdowns and more consistency between administrations across the UK. Our letters to producers published on 1 December 2020⁶ and 20 January 2021⁷ outline our expectations in relation to these statistics.

The improvements cover four broad aspects of vaccinations data. First, there should be more granular data on percentage take up, for example by, age band, ethnic group and by Joint Committee on Vaccination and Immunisation (JCVI) priority groups.

Second, in terms of UK-level consistency, the data available for each administration varies:

- Information on the percentages of individuals in each priority group that have been vaccinated to date is available for both Scotland and Wales.
- In England, breakdowns including ethnicity and some age bands are published by NHS England. However, it is disappointing that the publication of Public Health

⁴<https://uksa.statisticsauthority.gov.uk/correspondence/sir-david-norgrove-letter-to-matt-hancock-regarding-covid-19-testing/>
<https://uksa.statisticsauthority.gov.uk/correspondence/sir-david-norgrove-response-to-matt-hancock-regarding-the-governments-covid-19-testing-data/>
<https://osr.statisticsauthority.gov.uk/correspondence/ed-humpherson-to-stephen-balchin-nhs-test-and-trace-statistics-england/>

⁵ <https://osr.statisticsauthority.gov.uk/news/uk-data-on-hospital-capacity-and-occupancy/>

⁶ <https://osr.statisticsauthority.gov.uk/correspondence/ed-humpherson-to-producers-of-health-related-statistics-across-the-uk-statistics-to-monitor-the-uks-covid-19-vaccination-programme/>

⁷ <https://osr.statisticsauthority.gov.uk/correspondence/ed-humpherson-to-producers-of-health-related-statistics-across-the-uk-number-of-covid-19-vaccinations-administered-in-the-uk/>

England's COVID-19 vaccine monitoring reports⁸, that might have been a vehicle for more granular information about the vaccination programme in England, has been halted with no explanation as to why it has stopped or when it may restart. We have called upon the statistics producers to be clear on the data currently available and when more data will be published.

- Northern Ireland data are included in the UK dashboard. However, the more granular data are not released in Northern Ireland itself in an orderly manner. We have written separately⁹ to the Department of Health (Northern Ireland), requesting that the data are published in an orderly and transparent way.

Third, on vaccine type, Scotland is the only administration to routinely publish data on vaccination type (Pfizer BioNTech or AstraZeneca). It is in the public interest to have this information for all parts of the UK, particularly in the context of the media coverage on vaccine efficacy and sustainability of supply.

Fourth, it would also be helpful to have better information on those who have been offered a vaccine and those who have taken them up. This would help with understanding the reasons why some people may not be taking up vaccines, for example whether it is refusals, access or because they have not actually received an invitation to have a vaccine.

In addition to the areas outlined above, there are gaps remaining in health related COVID-19 data. For example, very little data exists on 'long COVID', despite emerging evidence that a proportion of people can suffer from symptoms that last for weeks or months after the infection has gone. We understand that the ONS is starting to look at this area¹⁰ and welcome this effort to fill an important data gap.

There are also gaps in the understanding of and information on variants of coronavirus. This is important in understanding the implications, for example whether data already collected on issues such as hospitalisations, deaths and efficacy of vaccinations are still applicable.

It is a fundamental role of official statistics and data to be used to hold government to account, and the lack of granularity and timeliness seen with some of the data makes it hard to do this. We recognise that producers are working intensively to make improvements and are keen to support these efforts.

Wider impacts

It is inevitable in such a fast-moving environment that there will be gaps in the data. To date the priority has been to fill the gaps most needed to understand the immediate and most direct impacts of the pandemic but at some stage the focus will need to shift from what is happening today to looking at what data are needed to fully understand the longer term consequences of the pandemic.

The issues are broad ranging and cover diverse areas such as:

- the impact of the pandemic on children and young people
- the long-term effects on health services
- the future health of the nation
- the impacts on inequalities

⁸ <https://www.gov.uk/government/publications/covid-19-vaccine-monitoring-reports>

⁹ <https://osr.statisticsauthority.gov.uk/correspondence/ed-humpherson-to-richard-pengelly-department-of-health-northern-ireland-number-of-covid-19-vaccinations-administered-in-the-uk/>

¹⁰ <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/datasets/prevalenceoflongcovidsymptomsandcovid19complications>

- the financial impacts of both the pandemic and the response to it

Answering society's questions about the pandemic must be done using both existing and new data. There are some longstanding statistical outputs on aspects of society that are likely to have been affected by the pandemic - for example, statistics on trade and migration. These outputs should be used to present analysis on the pandemic's impact on those specific issues. However, where existing data is collected through a survey, the data may be impacted by the difficulties in collecting face to face survey data over the past 12 months. It is also important to consider sooner rather than later where new data may be required.

2. Can you give a view on how well (or otherwise) the Government has communicated data on the spread of the virus and other metrics needed to inform its response to the pandemic?

During the pandemic we have highlighted two main areas of concern in the way governments have communicated data: accessibility and transparency.

Accessibility

Early in the pandemic there were lots of new sources of data being published by a range of organisations to support understanding of the pandemic. Much of these data were put out without narrative or sufficient explanation of limitations and could be hard to find. This made it difficult for members of the public to navigate the data and understand the key messages, which in turn could undermine the data and confidence in the decisions made on the basis of the data.

There have been improvements. For example:

- Data have been more effectively drawn together in dashboards and summary articles, such as the UK coronavirus dashboard and the continually evolving dashboards in Scotland, Wales and Northern Ireland.
- Publications have been developed to include better metadata and explanation of sources and limitations, and how the data link with other sources. An early example was the Department for Transport's (DfT) statistics on transport usage to support understanding of the public response to the pandemic. We also saw the Department of Health and Social Care (DHSC) develop its weekly test and trace statistics to include information on future development plans and in October an article was published comparing methods used in the COVID-19 Infection Survey and NHS Test and Trace.

However, it can still be hard to know what is available and where to find data on the range of issues people are interested in.

Transparency

Throughout the pandemic we have been calling for greater transparency of data related to COVID-19. Early in the pandemic we published a statement highlighting the important of transparency¹¹. We also published a statement¹² and blog¹³ on 5 November 2020. One of the prompts for this was the press conference on 31 October announcing the month-long

¹¹ <https://osr.statisticsauthority.gov.uk/news/covid-19-production-and-use-of-management-information-by-government-and-other-official-bodies/>

¹² <https://osr.statisticsauthority.gov.uk/news/osr-statement-regarding-transparency-of-data-related-to-covid-19/>

¹³ <https://osr.statisticsauthority.gov.uk/why-trust-and-transparency-are-vital-in-a-pandemic/>

lockdown for England. The slides presented at this conference were difficult to understand and contained errors, and the data underpinning the decision to lock down were not published for several days after the press conference.

We continue to see instances of data being quoted publicly that are not in the public domain. Most recently, for example, at the Downing Street briefing on 3 February, the Prime Minister said:

"...we have today passed the milestone of 10 million vaccinations in the United Kingdom including almost 90% of those aged 75 and over in England and every eligible person in a care home..."

At the time this statement was made these figures were not published. Breakdowns by age are not published for the UK as a whole. On 4 February NHS England included additional age breakdowns in its publication¹⁴ for data up to 31 January, including percentage 75-79 for the first time (82.6% having had a first dose), and 80 and over (88.1% having had a first dose). The 10 million first doses figure for the UK was reached on 2 February (published 3 February). Our view is that it is a poor data practice to announce figures selectively from a dataset without publishing the underlying data.

Our recent report on statistical leadership¹⁵ highlights the importance of governments across the UK showing statistical leadership to ensure the right data and analysis exist, that they are used at the right time to inform decisions, and that they are communicated clearly and transparently in a way that supports confidence in the data and decisions made on the basis of them. It sets out recommendations to support governments in demonstrating statistical leadership from the most junior analysts producing statistics to the most senior Ministers quoting statistics in parliaments and the media.

We will continue to copy letters to PACAC when we write publicly on transparency issues.

3. Can you give a view on how well the UK's statistical infrastructure has held up to the challenge of the pandemic? Are there key systems or infrastructure issues that need to be addressed going forward?

By and large the statistical infrastructure has been extraordinarily resilient. It moved from a largely office-based operation to being almost completely remote in a very short space of time. There were also fast adjustments to allow surveys and data operations for major household, economic and business statistics to continue in some form. Furthermore, new important statistical surveys were launched in a few weeks that would have taken months of planning pre-pandemic, while existing surveys were adapted swiftly from face-to-face to online.

The statistical system has also responded quickly to the need to provide more timely data, for example to support daily briefings.

Producers have been more open to creative solutions and new data sources, for example web-scraped data and PAYE Real Time Information. There have also been greater instances of data sharing. For example, the ONS coronavirus insights dashboard seems to be working towards being a 'one-stop shop' for several different producers and the devolved administrations to collate data in a single place. This appears to be encouraging communication between producers and is improving each week.

¹⁴ <https://www.england.nhs.uk/statistics/statistical-work-areas/covid-19-vaccinations/>

¹⁵ <https://osr.statisticsauthority.gov.uk/publication/statistical-leadership-making-analytical-insight-count/>

There are key infrastructure opportunities now that can be exploited, and it is important to question what elements of the new approaches should remain and which should change back to how things used to be done. For example, when and how best should data collection return to face to face households surveys? Should legacy surveys like the Annual Business Survey continue or should there be a move to new platforms or administrative data? How can the new data sources that have now come on stream be exploited even more? Is there a case for synthetic data to enhance existing data to help phase out large and expensive surveys? Can new survey platforms be used to answer short-term questions to help manage the impacts of the pandemic?

It is also important to learn lessons from mistakes that have occurred, at whatever stage of the process they occurred. One high profile example involved an error with testing data which meant there were delays to some data being included in the daily figures compiled by PHE¹⁶.

Errors like this reflect the underlying data and process infrastructure. OSR is currently exploring the use of reproducible analytical pipelines in government¹⁷. We are focusing on what enables departments to successfully implement RAP and what issues prevent producers either implementing RAP fully or applying elements of it. This work will give a further insight into infrastructure challenges and where improvements may be needed

Much of the data that are published are also drawn from administrative sources. In terms of monitoring the pandemic this has presented specific challenges. Public health, social care and hospital administrative systems are not connected to one another, which makes it time consuming to collate the data and puts quality at risk. Updating the IT infrastructure and data governance to make it possible to share information in a timely way is vital.

The pandemic has again highlighted the importance of analysts being involved in the development of operational systems, to make sure they are set up in a way that can best support data and evaluation needs.

The pandemic has also highlighted some of the strengths and limitations of the UK Statistical System. For example, analysts embedded in policy departments and devolved administrations have been able to quickly respond to emerging issues, but this has also been balanced with the complexity of having multiple organisations working on overlapping areas.

IT infrastructure as well as how statisticians organise themselves within and across organisations are covered further in our Statistical Leadership report¹⁸.

4. What should the Government learn from the issues with testing data that can be applied to vaccine data?

A key issue that has persisted throughout the pandemic has been the need for timely data to be published as quickly as possible. We saw in the early days of testing that the data were disorderly and confused. We were keen that this experience was not repeated with the roll-out of vaccines. So before the start of the vaccine roll-out, we wrote to producers of health-

¹⁶ <https://www.gov.uk/government/news/phe-statement-on-delayed-reporting-of-covid-19-cases>

¹⁷ <https://osr.statisticsauthority.gov.uk/systemic-review-outline-reproducible-analytical-pipelines/>

¹⁸ <https://osr.statisticsauthority.gov.uk/publication/statistical-leadership-making-analytical-insight-count/>

related statistics across the UK¹⁹, outlining our expectations that they should work proactively to develop statistics on the programme.

Drawing on the lessons to be learnt from the development of test and trace statistics, we outlined the following key requirements for vaccination statistics:

- Any publications of statistics should be clear on the extent to which they can provide insight into the efficacy of the vaccinations, or whether they are solely focused on operational aspects of programme delivery.
- Data definitions must be clear at the start to ensure public confidence in the data is maintained.
- Where statistics are reported in relation to any target, the definition of the target and any related statistics should be clear.
- The statistics should be released in a transparent and orderly way, under the guidance of the Chief Statistician/Head of Profession for Statistics.
- Some thought needs to be applied as to the timeliness, for example, daily or weekly data, to meet the needs of a range of users.

Encouragingly, we have seen signs that producers have learnt from their previous experiences. For the most part they made initial vaccination data available quickly; at first the numbers were fairly crude but they have continued to develop the data as time has gone on. We have seen also that whereas data were initially published on a weekly basis, producers very quickly moved to publishing daily figures with more detailed breakdowns provided in the weekly updates. This in part reflects a greater acknowledgement of the need to publish the data so that it can be quoted in parliaments and media without undermining confidence.

This is pleasing but the statistics remain in development and there is more to be done, as we have outlined in this annex and in our letter of 20 January²⁰. We also asked producers to publish development plans for these statistics and indicate if some data cannot be provided, as this will help users to understand the limitations of the data available.

**Office for Statistics Regulation
February 2021**

¹⁹<https://osr.statisticsauthority.gov.uk/correspondence/ed-humpherson-to-producers-of-health-related-statistics-across-the-uk-statistics-to-monitor-the-uks-covid-19-vaccination-programme/>

²⁰<https://osr.statisticsauthority.gov.uk/correspondence/ed-humpherson-to-producers-of-health-related-statistics-across-the-uk-number-of-covid-19-vaccinations-administered-in-the-uk/>