ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

EVIDENCE AND POLICY-MAKING

Work-in-progress report based on RSS event taking place on 21 June 2022

Introduction

This report documents the discussion at the third of the RSS's Covid-19 evidence sessions.[1] Before the session, we identified five questions designed to discover what lessons could be learnt around the role of statistical evidence in policy-making – both so that we can learn from what went well, as well as reflecting on where there are areas for improvement. We sought to bring in a wide range of views during the discussion at the event, and these are reported below. However, even over a two-hour meeting, only so many people were able to speak, so this is intended as a reflection of views expressed during the meeting and should not be read as representing the views of the RSS.

The five groups of questions we explored were:

1. What is the value of statistical thinking in a national emergency? Are there both positive and negative examples from the pandemic of times when a statistical perspective improved or could have improved evidence?
2. How can statistical modelling be used in epidemics? How should such modelling be best communicated to policy-makers?
3. How should statisticians be involved as experts when there is an emergency? How should statisticians prepare and train for this situation?
4. How can statisticians give balanced advice to policy-makers and be trusted as an intermediary?
5. How did global data feed into the UK's decision-making process? What lessons should we learn on international data sharing going forward?

List of speakers

Main speakers

- **Sarah Walker** – Professor of Medical Statistics and Epidemiology, Nuffield Department of Medicine
- **Sebastian Funk** – Professor of Infectious Disease Dynamics, LSHTM
- **Emma Rourke** – Director of Health Analysis and Pandemic Insight at ONS
- **Nick Jewell** – Chair of Biostatistics and Epidemiology, LSHTM and Professor of Biostatistics and Statistics at the School of Public Health at the University of California, Berkeley.

Contributing speakers

- **Roger Pielke** – Professor of Environmental Studies, University of Colorado
- **Sylvia Richardson** – RSS President, co-chair of RSS Covid-19 Task Force and Programme Leader, MRC Biostatistics Unit, University of Cambridge
- **Rob Harrison** – Director General Analysis, Cabinet Office
- **Aditya Goenka** – Professor in Economics, University of Birmingham
- **Jen Persson** – Director and Founder, Defend Digital Me
- **Ian McKendrick** – Head of Consultancy, Biomathematics and Statistics Scotland
- **Thomas House** – Professor of Mathematical Sciences, University of Manchester

[1] This document is a work in progress – a final version will be published in 2023. If you notice errors or omissions please email policy@rss.org.uk.

- **Harrison Schramm** – Principal Research Scientist, Group W
- **Andrea Rehman** – Chair of RSS International Development Section and Assistant Professor in Medical Statistics, LSHTM
- **John Aston** – Harding Professor of Statistics in Public Life, University of Cambridge.

1. What is the value of statistical thinking in a national emergency?

Sarah Walker discussed her personal experience as chief investigator and academic lead for the Covid-19 Infection Survey. The central reason that statistical thinking is important, she explained, is because national emergencies rely on information to inform a policy response, and information requires data analysis or new studies. Both require statistical thinking because the statistical question being addressed should determine the design of each analysis or study, and because avoiding bias is crucial. “You can’t fix by analysis what you bungled by design”.

It isn’t just about design, **Walker** argued. A statistician’s responsibility goes beyond just making plans – they must seek assurance of cooperation, and maintain constant touch with the work as well as the interpretation of results.

In the context of a pandemic in particular, **Walker** suggested, there is an interdependence between statistics, operations (what goes on in the field) and the underlying epidemiology. If you lose any of the three you end up with a disaster – so it is not about statistical thinking in isolation, in must be in connection with these other aspects.

She highlighted the PPATI Study (Post Positive PCR Antibody testing initiative) as an example of where this went wrong. This was a study that did two antibody tests in a number of people with positive PCR tests in the national testing programme. The first antibody test was taken within six days of getting the positive PCR test result and the second three weeks later. In the information sheet it was stated that doing an antibody test after a PCR result helps the NHS learn how likely people were to get Covid again and how the body’s immune system responds to the virus or vaccine. This demonstrated a critical failure to understand or acknowledge that six days after you get a positive PCR result is almost always already too late (as people have symptoms before they test positive) and a failure to understand that just testing PCR positives doesn’t tell you about risk – it is like testing dead people and thinking that tells you about the risk of dying, **Walker** argued.

In terms of where statistical thinking hasn’t helped us so much, is the concept of “what is good enough” and not letting the best be the enemy of the good. As statisticians the avoidance of bias is incredibly important. But the pandemic has made **Walker** question what level of evidence is sufficient – or how much bias is acceptable in terms of decision-making – particularly when you’re trading it off against other aspects. In a pandemic timeliness is crucial. The Test Negative Case Control (TNCC) designs for vaccine effectiveness are pertinent here. These compare vaccination status in symptomatic people presenting for testing and testing positive, versus testing negative. There were substantial and important concerns about how testing behaviour and failure to adjust for confounders relating to health status may have led to bias. However, even if this was a little bit biased, it will not necessarily have that much impact on decision-making.

Walker agued that increasingly, it is important to think about what is a “good enough” analysis, if it is presented with a bit of information about plausible bias. And thinking about how to do this is a key challenge for the profession. The challenge is deciding what is and isn’t good enough – which requires balance and judgement. A good area to look at to understand the trade-offs involved is incidence estimates. She argued that the testing data was good enough to identify trends even though we know it is biased by testing behaviour. The estimates produced from the Covid Infection Survey are more accurate but are not timely because they use prevalence to estimate incidence and there is an inherent delay. The ZOE estimates use the incidence of new symptoms plus positivity estimates and a large number of underlying assumptions. Where is the balance between bias and “good enough”?

There is another issue around the precautionary principle – the idea that decision-makers might adopt precautionary measures when evidence is uncertain and stakes are high. For example, considering the decision whether to advocate facemasks at the beginning of the pandemic – how much evidence is actually needed in terms of deciding what is good enough, especially when there may be a precautionary principle?

Sebastian Funk also highlighted some examples that showed the value of statistical thinking during the pandemic. He highlighted three key areas:

1. Respect for data and an understanding of its origins and especially its limitations.
2. Transparency of methodology, so that observed trends and observed results can be explained clearly and investigated.
3. Statistical interpretation of results – whether that's via credible intervals or effect sizes – and the corresponding uncertainty.

Many of the statistical tasks (eg, the five examples that **Funk** highlights in his discussion of modelling, covered in §2) are quite common. The example below is a recent one from the UK Health Security Agency (UKHSA) and is a chart showing the incidence of monkeypox by day of symptom onset.

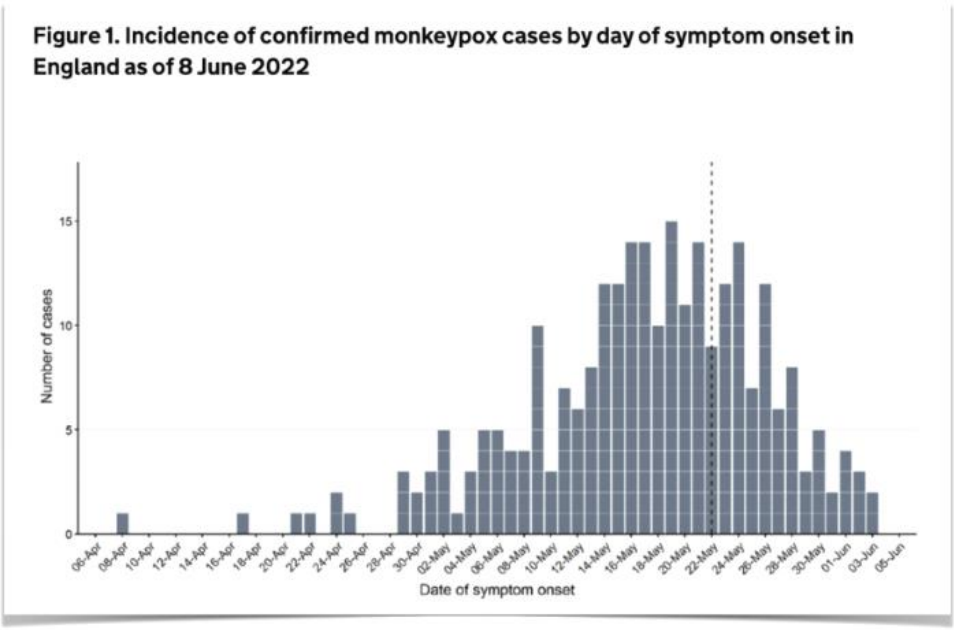


Figure 1. Incidence of confirmed monkeypox cases by day of symptom onset in England as of 8 June 2022

UKHSA, 2022

There are statistical methods that would enable this chart to be presented as a nowcast to give a clearer indication of whether cases are declining or whether they might be plateauing or even increasing. There are two main issues with public health agencies around the world needing to perform this task. First, it needs to be done quickly – generally what tends to be "good enough" tends to be what is available or possible on a very short timescale. This means that methods – and people with the skills to use them – need to be ready continuously. Any method that is not well-implemented in software is essentially useless in a public health context.

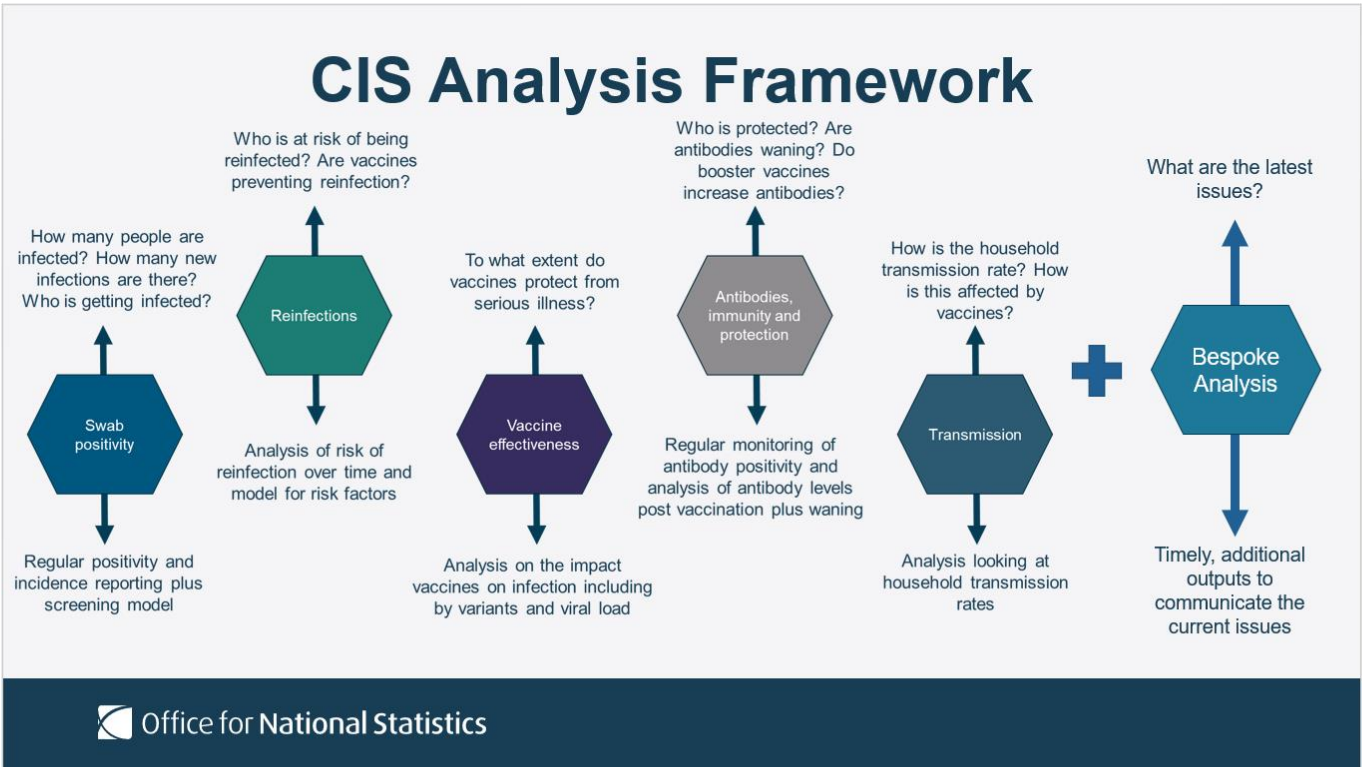
Second, plenty of methods for nowcasting have been published, but much less has been published on their relative performance against each other, or relative strengths and weaknesses. **Funk** highlighted one example of a project that is trying to address this, the [German Covid-19 nowcast hub] (as prospective evaluation, COVID-19 hospitalisation data in Germany suffers from exactly the same problems as we see in the UK).

The importance of statistical methods during a pandemic – and the importance of having these ready to deploy – means that there is a huge opportunity for statisticians now to get involved in collaborative efforts to improve preparedness.

Emma Rourke set out the ways in which the Office for National Statistics (ONS) has supported the UK's pandemic response and the lessons that can be drawn about the value of statistical thinking. Statistics as a subject is inherently both disciplined and deeply creative. Our ability to bring together a range of different data and apply techniques to provide insight and inform and enlighten the policy discourse is integral to good policy-making. Discipline and rigour are important qualities, but statisticians also need to engage with decision-makers and use their creativity with data to help answer pressing policy questions. This has been a key feature of how the ONS has been operating through the pandemic.

Statistics, **Rourke** said, also permits a coherent analytical discourse. Quite often the way in which analysis develops is very discursive – a piece of analysis is run, discussed, explored, hypotheses to explain it are developed and the process is repeated iteratively. In the context of a national emergency, this means doing things that are "quick and dirty" and being comfortable with what is good enough. How statisticians communicate all this with non-statisticians/analysts is important. This involves supporting them to become more confident and competent, to understand statistical techniques and where the data has come from, and to be able to present uncertainty in words or visualisations to help other colleagues to engage with the data.

One way that statisticians can work with policy-makers to support the navigation of all of those different threads is by setting out an analysis framework. Below is an example of how the ONS has set one out using the Covid Infection Survey, allowing analytical discourse and engagement with policy colleagues to continue to iterate the questions – the question at the start may differ to that at the end. The partnership through the flow of the conversation is critical.



The same is true for visualisation. Complex messages can be difficult to communicate, and a lot of the ways in which we might visualise data can be really quite challenging for non-analysts. The UK has come some way in analytical literacy both among the public and in policy-makers. The ONS has done a lot of work to iterate the way in which data is visualised to ensure that they meet people where they are. The important thing is that this visualisation is understood – it can't be helpful if it's not understood. This is especially important for helping the public to be part of the wider policy discourse.

Nick Jewell gave some examples of where statistics has helped inform decision-making and where greater involvement of statisticians may have helped. The ongoing surveillance data provided by ONS and REACT was incredibly valuable – the trends identified by those studies (probably more than individual point data) were very useful in guiding policy. There did, though, seem to be an issue with interoperability – ie, the early response perhaps indicated a need for interoperability requirements for data to allow very rapid sharing across jurisdictions. There were examples in the UK of this working well – eg, with sequencing data giving us the ability to detect new variants quickly and to estimate their prevalence. But there needs to be greater emphasis on interoperability across the piece.

Jewell also pointed to experimentation as a way in which statistical thinking can improve pandemic responses. This worked well with vaccine development: systems were implemented very quickly for the RECOVERY trial platform. This is the type of situation that statisticians are very well trained and prepared for. While statisticians were vocal and effective on the need for experimentation and getting information about the effectiveness and safety of vaccines and therapies, we've been less vocal about the need for experimentation for other mitigation measures that are more socially oriented – including school closures, mask wearing, quarantine rules etc. If we had the mechanisms set up and had considered how this might have been done in advance, it would have been possible to provide more evidence about the effectiveness of a lot of these measures. **Jewell** argued that we just didn't try to experiment enough. Experiments in a social context are perhaps not as easy statistically as vaccine trials, but they are still doable.

Jewell also identified an overreliance on ecological comparisons of infection data and hospitalisation data. Both of these are collected poorly and when they are used to address causal questions as to why one country or one region is doing better than another, there is a lot of "statistical nonsense" we would never depend on to test therapies.

John Astin made the point that what is "good enough" very much depends on the phase of the pandemic – there will be a lower threshold for this at the start than later on as the pandemic continues and more information is available. **Walker** also emphasised this – when you look at our response to the Alpha variant, we were a dollar short and a day late in many ways. But by the time Delta and Omicron arrived we were able to move much more quickly. With Omicron in particular we had rapid estimates of the impact on hospitalisation and vaccine effectiveness.

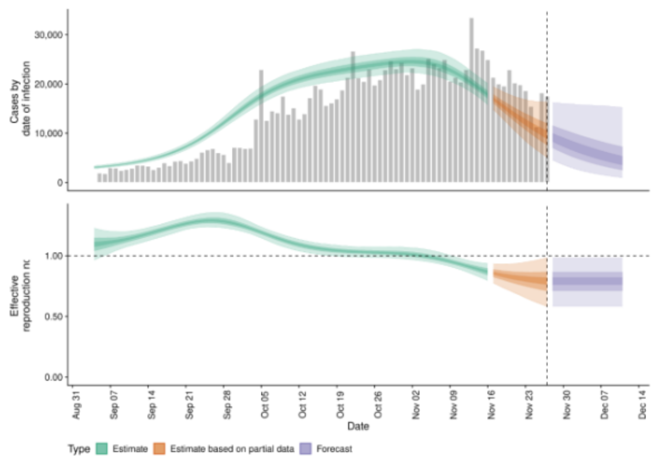
Jen Persson gave an example of how statistical thinking could have better informed another area of pandemic-related policy: the measures used to support children with remote learning when schools were closed. In the 2020 rapid response to remote learning, the Department for Education told schools that the number of devices available, that is to say, laptops, routers and hardware, would only be confirmed at the time of ordering based on stock availability. Schools were reportedly able to only claim 20% of what they expected, and it's important to look at what assumptions and data were used in calculations. Her organisation, through FOI requests, has found evidence that suggests that the DfE assumed a) that private device needs would be met by some extent of existing laptops held and b) that the free school meals allocation did not give an accurate picture of device requirements, and these two influenced their distribution assumptions. They also made the assumption that the first devices sent to schools would go to teachers and so they removed devices equivalent to that percentage of teachers in an average school from the average number of devices that they already held. A key lesson here is that a better proxy for poverty than free school meals is required. It would be interesting to look how equitable the distribution of technology was by factoring in the areas of no recourse to public funds and free school meal eligibility by ethnicity.

2. How should statistical modelling be used and communicated?

Sebastian Funk gave five examples of ways in which statistical modelling played a key role during the pandemic. First was the estimation of the R number – the most common method is the renewal equation model, where you have the number of infections, which is the sum over the distribution of generation intervals (the time between infection and infecting others) multiplied by infections at a previous time and multiplied by the scale factor. This is a common type of model (an autoregressive time series model) where the scaling factors over time are given and there is a common scaling factor at each point which is being estimated. There is a complication in the case of Covid, which is that we don't observe infections directly – so there is statistical work to be done to estimate the number of cases. But all of this can be done with standard statistical techniques. One example of how this could be done, using the modelling language STAN, is shown below:

Example: Estimating the reproduction number

$$I(t) = R_t \sum_{\tau} g_{\tau} I_{t-\tau}$$



Abbott et al., *Wellcome Open Res*, 2020

This gives you an estimate of the reproduction number over time. This was one of the models that fed into the UK government's official estimate of R.

Second, **Funk** highlighted work done on the transmissibility of new variants. Below is an example of work done on the Alpha variant:

Example: Transmissibility of new variatns

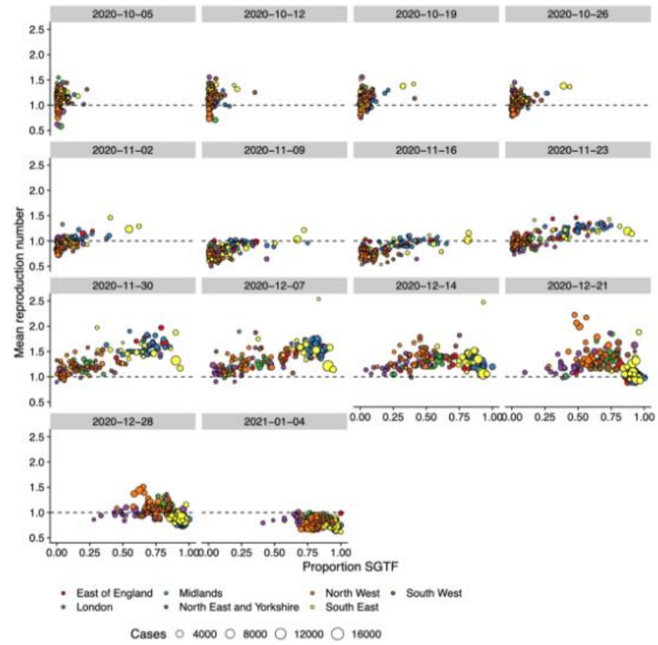


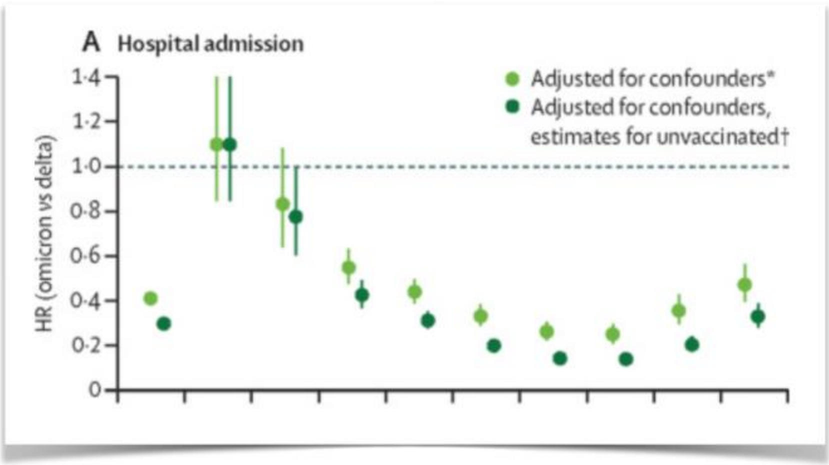
Figure 2: Mean reproduction numbers using a generation time with a mean of 5.5 days since the week beginning 14 September, 2020, compared to the proportion of all test-positives tested for S-gene that tested S-gene positive/negative that week. Each point represents one UTLA, with the size given by the number of cases in the week following the week of the given reproduction number to account for the delay from infection to testing.

Davies et al., *Science*, 2021

This is work done on the Alpha variant that emerged in winter 2020. Here, looking at local reproduction numbers over time (week-by-week), we can see that as the proportion of people in each local area with the Alpha variant increased, we saw an increase in the reproduction number. We can use that to estimate the advantage in terms of reproduction Alpha had over the previously circulating variant.

Third, as new variants don't just have different transmissibility – they also have different severity – it is important to also be able to model their severity. An example of this is below:

Example: Severity of new variants



Nyberg et al., *Lancet*, 2022

This example is for Omicron, compared to Delta. This was one of the pieces of evidence showing that the hazard ratio of hospital admissions from Omicron was quite a lot lower compared to Delta when adjusting for confounders, which is why the Omicron wave had much of a lower impact in terms of hospital admissions and deaths. This was calculated using a common statistical technique – Cox Proportional Hazard Regression.

The fourth example of statistical modelling that **Funk** gave was nowcasting. Below is an example – the nowcasts of deaths by the UK Health Security Agency and the MRC Biostatistics Unit in Cambridge.

Example: Nowcasting deaths

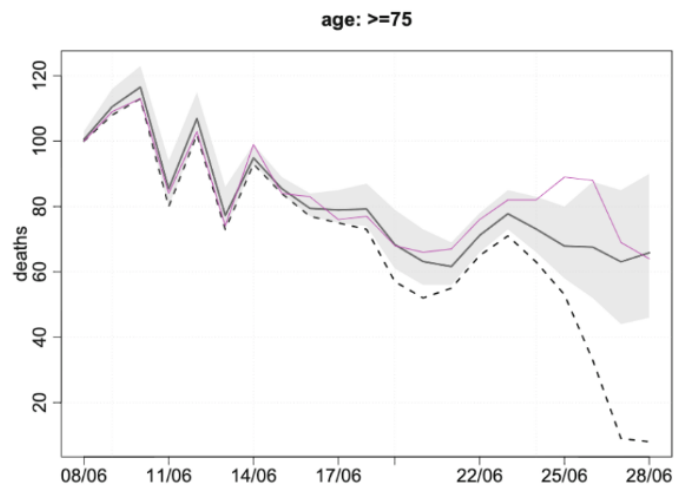


FIGURE 4 Estimated numbers of deaths occurring on each day (black line) in the ≥75 stratum, with corresponding numbers of deaths reported by 29th June (broken line) and the true numbers (purple line). Posterior 95% credible intervals are shown by shaded region. For ease of viewing, only most recent 21 days are shown.

Seaman et al., *J R Stat Soc C*, 2022

A really common problem in epidemiology is that quantities are observed with a delay. The observations above are deaths as of the 29th of June. Because these deaths are recorded with a delay, they always fall off towards the present (the dashed line). If these delays are understood or known from previous iterations of the data, this can be corrected for to produce a nowcast such as that shown by the black line (with shading for uncertainty). The purple line shows the retrospective view of what really happened.

A lot of the public focus on statistical modelling was on the scenario models that came out of SPI-M. **Funk's** fifth example was the calibration of these scenario models. An example is shown below:

Example: Calibrating scenario models

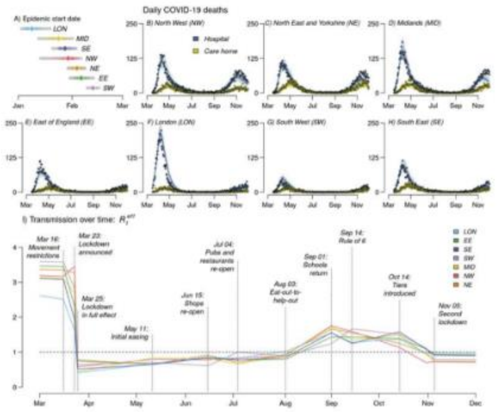


*Figure 1: Trajectory of the England COVID-19 epidemic. **A,** The inferred epidemic start date in each NHS England region. **B-H,** The model fit to reported daily deaths from COVID-19 in care homes and hospitals for each NHS England region. The points show the daily data, solid lines the median posterior and the shaded area shows the 95% CrI. **I,** The mean effective reproduction number within the general community (i.e. excluding care homes) in each region from March to December. Vertical lines and labels represent dates of key policy changes, defining the breaking points of the underlying piecewise linear transmission rate. Dashed horizontal line depicts reproduction number of 1.*

Knock et al., *Sci Trans Med*, 2021

Here two scenario models were calibrated using Markov chain Monte Carlo statistical techniques. This is just one of the examples where this was done, and the image above shows how this was used to model different scenarios and understand their potential impact.

Nick Jewell argued that, while these are beautiful examples of how modelling can help us understand what's going on during an outbreak – they're not a panacea. Without data, even the most sophisticated models won't be much help. **Jewell** argued that reliable population data is more informative and better in translating information to the public than complex mathematical models.

Jewell also highlighted a number of limitations in how forecasting models should be used. The use of the reproductive number as a policy trigger is insufficient. This is because the estimation of the number is often crucially delayed from the timing of infection events and based on incomplete surveillance information. Proxies of the reproductive rate are also complex and poorly understood – the reproductive rate is itself a complex measure and depends on several different underlying phenomenon. Models are effective in helping to retrospectively understanding epidemiology, but predictions are usually inaccurate (beyond reported uncertainty). Focusing on predictions as the output of models has not been very useful and has risked undermining the use of modelling in general. The most effective use of models lies in ranking and assessing potential mitigation strategies.

3. The role of statisticians as experts during a crisis – how should we prepare?

Sarah Walker suggested a couple of key things to focus on. First, a breadth of experience is important. When you've got an unknown unknown, it helps to have seen as many similar situations as possible. Second, it's important to value diversity in statistical approaches: a different approach is not necessarily wrong. That means diversity in design – this is why it was beneficial to have both the Covid Infection Survey and the REACT survey. They were very different designs: the infection survey was a household survey, selected from address lists, and was longitudinal, delivered by study workers and post-stratified to account for sample characteristics; REACT was an individual survey, selected from GP lists, was repeated cross-sectionally, used postal kits, online questionnaires and samples returned by post, and was weighted to account for sample characteristics.

Diversity in analysis is also important. Deciding which models best correspond to the real world situation is difficult – multiple analyses help build understanding. A good example of this was vaccine effectiveness analysis: there was UKHSA's test negative case control approach as well as approaches that used Poisson regression (Neil Ferguson at Imperial), stratified Cox proportional hazards regression for time from positive test to hospitalisation (De Angelis, Presanis and Nyberg at Cambridge) and OpenSAFELY's "target trial" approach. These different approaches make different trade-offs (different speeds and different potential biases).

There is also a big piece about not forgetting the basics – especially confounding and interaction. The analysis comparing Omicron and Delta hospitalisations showed the critical importance of adjusting for age very finely (given background shifts in respiratory infections) as well as adjusting for calendar date, deprivation, local area effects and so on. Statisticians should remember that without data linkage you don't have good information on prior health conditions and you can't adjust for that properly.

It is important for statisticians to value how data are collected – and not just the data themselves. For linked data, understanding the provenance of the data becomes even more important. This is something that we need to train statisticians to think more about. Who gets tested where? What database is that test result in? Because not being positive doesn't mean negative. There are issues with coding bias – there's a difference between hospitalisations with and from Covid that need to be appreciated.

Emma Rourke drew on her experience at the ONS to set out some thoughts on how statisticians should be involved as experts during a crisis. She argued that it was important for statisticians to aim to be an agile partner – this means making sure to be there right from the beginning, fully embedded in the policy discussions. It is important for this to be meaningful and for statisticians not to just be add-ons. The ONS managed to launch the Covid Infection Survey, moving from conception to first visit, in just ten days, which is agile for a government body – but it is not just about launching it: it also needed to constantly move and develop to ensure that it was able to provide the evidence that policy-makers needed and so that statisticians could provide their expertise at the right moment.

The multiagency approach of statisticians is also important. The ONS's partnership with UKHSA has been integral to their ability to ask the right questions. Statisticians need to feel like they're part of an orchestra: they don't always have to be the lead violinist, but they are a critical part of the overall symphony. This means statisticians embedding themselves in public discourse and being really approachable. Statisticians need to find ways to communicate with the wider multidisciplinary community – Sarah Caul (one of the ONS's tweeting statisticians) provides a good example of how this can be done well. Transparency and openness to challenge helps ensure that the right analysis is developed as quickly and helpfully as possible.

The pandemic has been a catalyst for government statisticians – but now that they have learned about what is needed, it is possible to prepare and train in the right way in preparation for future crises. For statisticians this means thinking about how to increase their agility and how to align themselves with key questions of the day – inserting themselves in multidisciplinary, multiagency conversations. We need the data and the infrastructure to enable preparedness which are really crucial for effective working. Having a toolkit is also important. This can't be one size fits all: we need a variety of toolkits, and the confidence, maturity and variety of experience to constantly evaluate and reflect on the effectiveness of a given approach.

This is all part of the objective of the Integrated Data Service that the ONS is currently developing – which aims to help this sort of data-driven innovation at scale.

Nick Jewell also highlighted the importance of having statisticians at the table – they are trained in experimentation, communicating uncertainty, synthesising evidence and they are used to working in interdisciplinary teams. Statisticians have been working effectively as part of decision-making teams in the pandemic in the UK and elsewhere. This should be seen as very much be part of the portfolio of a statistician and their training should reflect that.

Ironically, **Jewell**, pointed out – at the start of a crisis decisions need to be made without a huge amount of evidence and data to support them. So, before and during the early stages of an outbreak, the role of statisticians has to be the development of effective data systems and surveillance systems. The UK has underinvested in surveillance systems for the natural reason that when there's not an immediate need and money needs to be saved, it is easy to cut this type of data collection process because it does not obviously have immediate value. But it is absolutely crucial to have strong data infrastructure as part of pandemic preparedness – so that early decisions can be as well-informed as possible. The UK, he argues, was not well-served in this respect at the start of the pandemic and a lot of subsequent effort has been made to address that.

When designing these systems it is important that statisticians, **Jewell** argued, think about not the best way to handle information for its own sake – but the best way that statistics can impact policy. These are different, but connected goals – as statistics can only best impact policy when it's based on accurate and informative surveillance data. Epidemiologists are trained on four key questions when a disease occurs, particularly a novel disease: why, who, when and where. Statistics helps provide designs for studies that can answer and address those questions. For a lot of the pandemic, especially in 2020, policymakers were flying blind because these questions were not well addressed by the UK's existing data systems.

Jewell suggested that rapid response training would be beneficial for statisticians and decision-makers. Earlier in his career, Jewell was involved in preparation for earthquakes in California. There's a very high risk of an earthquake over an extended timeframe. When those rare events happen it's an immediate crisis, with an immediate response needed. People need to be trained how to do that. So they would do training sessions where policymakers were put in a room, and statisticians would feed them data – usually bad data indicating bad outcomes – and help them learn how to make policy decisions and make trade-offs.

Statisticians, as far as possible, should also separate themselves from ideology. They should try to represent data and information in a way that's ideology-free, and not be too influenced by their own idiosyncratic opinions.

Jewell suggested there is a role for statisticians both in helping to prepare appropriate stockpiles as part of pandemic preparedness, and in specifying the scientific technology that should be used in testing. Most statisticians tend to be involved in research agencies – ie, in researching questions from an academic or a statistical point of view. But there's a great need for statisticians to become involved in mission agencies: helping governments and other local jurisdictions implement research ideas into the field.

Walker and **Jewell** both made the point that statisticians need to prepare to understand what is good enough to inform decision-making during a crisis. Part of how that can be understood is through experience – but it is important for experienced statisticians who have been through a crisis to pass on that experience to new statisticians. There is also a point about experimentation here: we don't really know what is good enough until we are able to test it. We don't really know if any mitigation strategy will really make an impact without testing it – sometimes things have been mandated for everyone because it doesn't seem harmful and might be beneficial, but that isn't conducive to gaining public trust. Learning what is good enough by experimentation could be useful because often our research about what might be good enough is based on the laboratory setting, which obviously has differences to the field. Sometimes things that work well enough in the lab don't in the real world – so experimentation has a crucial role to play in helping to understand what's good enough.

Sebastian Funk emphasised this point – at the time that analyses are made, statisticians know that they are doing things that involve biases – but we don't know the scale of the biases at the time. Usually, a model or analysis is needed rapidly and people do their best, but there's not an objective standard that we can then apply as to whether that best is good enough. **Walker** suggested though, that if we have an ensemble of models all of which seem to be moving in the same direction then, whatever the biases are, that can give you confidence.

There was some discussion about the role that the wider statistical community can play during a crisis – in addition to statisticians who are already in government, could those statisticians in the wider community (eg, in academic settings) have done more to support the response? **Walker** argued that the community mobilised itself fairly effectively – she saw statisticians from a wide range of backgrounds contribute an enormous amount on a regular basis. **Funk** suggested that, while there has been great sharing of insights, results and methods, in terms of tools or steps of analysis there has been a lot of duplication and a lot of different groups saying basically the same thing due to a lack of sharing. **Thomas House** also made this point: there aren't the incentives within academia to promote collaboration on this type of thing. The research community is open-spirited and will exchange ideas, but in terms of practical collaboration and actually helping each other, more could have been done. **Funk** agreed with this – junior researchers are incentivised to prioritise writing papers for high impact journals, because that is what their success will be judged on. Writing code that others can use would be very public-spirited, but it won't be rewarded in career progression.

Aditya Goenka gave an economist's perspective on the role of evidence and policy-making during the pandemic. There was considerable interest in using the data on infection and mortality, as well as on the impact of non-pharmaceutical interventions on the economy and society. However, it was not well appreciated that the variables in indicators were jointly determined, so drawing any inference on causality was problematic. He gave the example of analysis of lockdowns, where the health versus wealth trade-off was of interest to policy-makers. Economics outcomes and policy response are determined jointly, and none of these are exogenous in a statistical sense. So, looking at the effect of a lockdown on economic or infection indicators can be misleading. School closures was another example: understanding the impact of closures and openings on infections was very important, but all schools were shut or open at the same time alongside other interventions. So there was no natural variation to shed light on the causal effect of a school opening or a shutdown. Economists use a variety of econometric techniques to make these causal inferences, to study the effect of policies on different variables. Some of these include natural experiments, randomised control trials, regression discontinuity analysis, instrumental variables, etc. These should all be part of the toolkit for statisticians engaging in policy development.

Iain McKendrick gave his perspective, based on working at EPIC (Epidemiology, Population health and Infectious disease Control), which has been funded by the Scottish Government for sixteen years to provide evidence-based advice to policy-makers on exotic animal disease outbreaks. EPIC is a virtual centre, employing staff from six partner organisations including statisticians, mathematical modelers, data managers, epidemiologists, social scientists, economists and science policy interface brokers. All of these roles are very important. The long-term funding and existence of EPIC is critical in providing key positive outcomes. There needs to be: staff being in post with the right balance of relevant skills; trusted working relationships within the group and with key policy staff; good understanding of the needs of the policy customer in hypothetical but likely situations; access to key datasets negotiated in advance, managed in a GDPR-compliant fashion, and actively curated for ongoing use; multiple analytical pipelines and model code reports ready in advance for use at short notice in an emergency (not just to automate things, but also to leave space for critical and creative thinking on the fly); and a realistic understanding of how hard all of this is. These positives don't happen spontaneously, they need long term funding and careful management.

Thomas House emphasised the need for diversity in input. That needs to be the right kind – our media environment tends to make it too easy for people with a simple and quite extreme view on something to become prominent and promote misleadingly simple answers or the view that a particular intervention is so terrible that it should never be considered (or conversely that some intervention would be a panacea). The solution to this is hard, but a lot of the real scientific debate was happening in a way that would have been great for the public to be more involved in: things like how rapid testing was or wasn't used, how general test and trace was used, and which kinds of non-pharmaceutical interventions represented the right balance. A lot of the debate in the media, conversely, focused on people who could give a quick sound bite. So, as professionals, it is important for us to think about how we can involve greater public engagement in scientific debates.

4. How can statisticians give balanced advice to policy makers and be trusted as an intermediary?

Sarah Walker emphasised that everything is done by people. If you look into social psychology it will tell you that the most important thing is that there are in-groups (a social group to which a person psychologically identifies as a member) and out-groups (a social group to which they do not identify). Fundamentally, statisticians may need to recognise that data and facts are not everything. To be trusted as an intermediary, an individual has to become part of the in-group (which statisticians have done a good job of during the pandemic – really getting involved in the Joint Biosecurity Centre (JBC) and UKHSA), or you've got to work with people who are part of the in-group. It is important that, instead of just relying on facts, statisticians need to be able to tell the story. If you want to give advice and be trusted, you've got to tell people what they need to hear in a way that they can understand – and this sometimes means relying on opinion, anecdote or metaphor in addition to facts. While statisticians may be 'facts people', not everyone is.

Roger Pielke spoke about how science and evidence can inform policy-making. He suggests that advice is best understood as guidance regarding a decision or a course of conduct. By definition, advice is in the context of someone choosing among different courses of action. Within advice there's different types of advice. Science advice is the most narrow. It can be focused on disciplinary or interdisciplinary expertise. Expert advice is broader. It can include practitioners. It can include multidisciplinary expertise. It's much more grounded in skill or knowledge than just science advice, which is based on knowledge. Political advice, which is really important in decision-making contexts, has to do with bargaining, negotiation and compromise: the essence of democratic politics.

Pielke offered a short summary of his book, which aimed to outline roles and responsibilities for providing scientific advice in policy settings. In the book there are four categories.

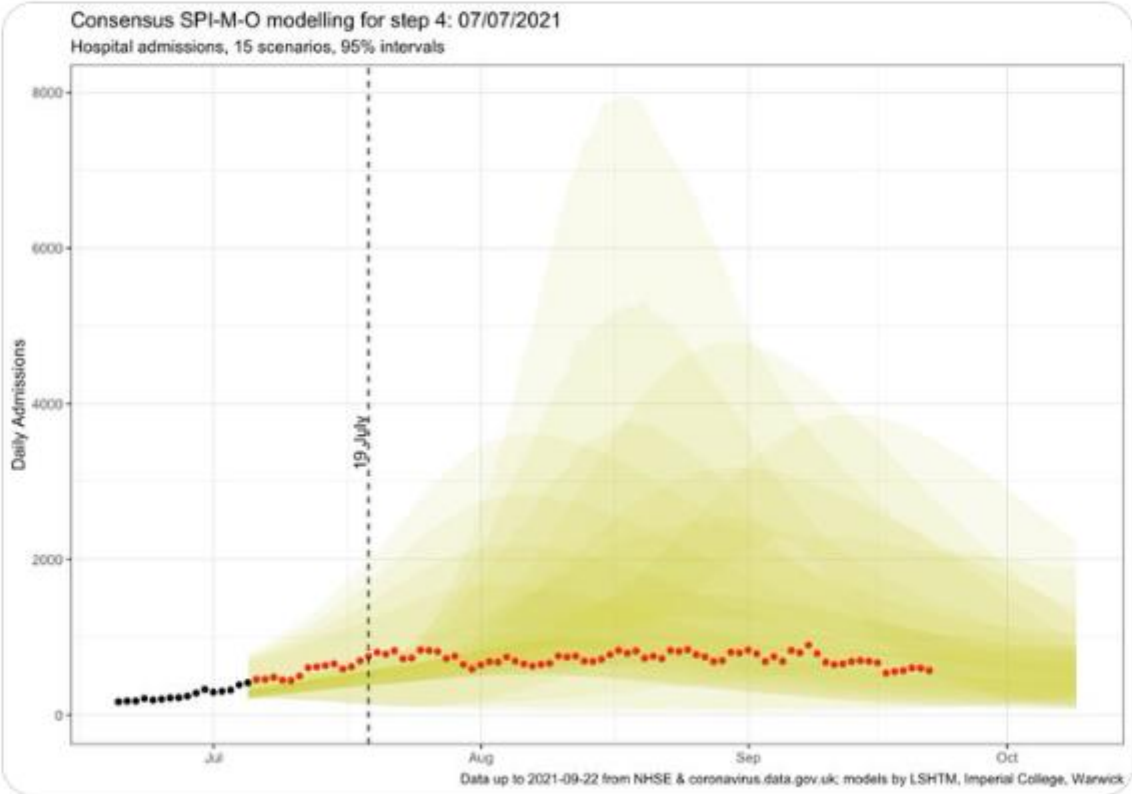
1. Pure scientists. We can debate whether pure scientists actually exist, but the defining characteristic is that they have no desire or interest to be connected to policy. In practice we find this is a very hard distinction to make because people often advocate that they are just pure scientists, but they are in fact trying to influence decision making.
2. The science arbiter. The science arbiter is characterised by a back-and-forth relationship with policymakers, where the policymaker will ask a question that can be answered with the tools of science experts, and scientists will try to answer the question.
3. The issue advocate. The defining characteristic of the issue advocate is a desire to try to reduce the scope of choice of policymakers, usually down to one preferred option. So, eg, people advocating for lockdowns or no lockdowns.
4. The honest broker. The goal of the honest broker is to expand or clarify the scope of options to empower decision makers to make choices. So think of the honest broker like a travel website. It tells you where you might go, how much you might pay, but it doesn't tell you where you have to go.

Sylvia Richardson argued that statisticians and data scientists need to add a voice to help clarify the distinction between scientific evidence and its consequences – we need to help policymakers understand the way which immediate and time-limited evidence can help inform decisions. Statistical expertise has a unique place in the sequence of events that starts with scientific findings and ends with a specific policy implementation. This includes: advising on the analysis and interpretation of data to support a given conclusion; identifying the need for new data to fill a knowledge gap; helping in designing data collection; highlighting sources of uncertainty and quantifying their impact; and assisting in evidence syntheses.

Many of the issues that confronted the UK have benefited from the inclusion of a statistical perspective right from the start. **Richardson** pointed to examples from other speakers that show where this went well. However, there were also missed opportunities, for example, evaluating the performance of the Test and Trace system for stopping transmission. The evaluation would have benefited from embedding agile data linkage right from the start, as well as regular reporting of epidemiologically interpretable quantities.

Rob Harrison gave some personal insights about what sort of advice had impact. There was always a high level of uncertainty and a very wide range of views about the best course of action. It was essential to provide decision makers with a single shared picture of the available evidence, as a common starting point for those discussions. So conversations should be in two parts – initially debating what the evidence can tell you, and the secondly considering what to do about it.

A persistent challenge was how to communicate complexity and uncertainty in ways that were accurate but quickly and clearly accessible to a lay audience. **Harrison** pointed to a couple of examples. First was the Covid infection survey, which provided a world-class, high-frequency, large-scale time series to track the trajectory of the pandemic, but whose headline results could be presented to decision makers in a single visualisation. The second is from Graham Medley's Twitter feed – he posted scenarios from multiple modelling groups, layered to show the shape of the plausible future space and the distribution of probabilities within it in a way that was much more helpful to policymakers than fixating on one or two (usually dramatic) scenarios.



Harrison also made a point about people. The UK is very fortunate to have a strong contingent of professional statisticians, and being able to draw on professionals from the academic community and beyond was essential to the government's response. Two key enablers of this were transparency and trust. On transparency, it was essential that the statistical reports from SAGE and SPI-M were all published. The UKHSA Covid dashboard deserves a special mention here, as it made a huge range of government data available in machine-readable format for others to use, which really improved not only the quality of analysis within government but also the quality of the public debate.

And finally, **Harrison** suggested that the impact of advice was greatest when policy was designed to be data driven. A really good example of this was the Spring 2021 road map. This was designed to allow a minimum of three weeks of data plus time for decision-making and implementation. This shows the benefit of statisticians and policy-makers working together to design policy so that statisticians can inform it in an ongoing way.

Harrison Schramm set out some lessons from his career at the interface of statistics and public policy in the United States. Considering the interaction between evidence-based decisions, which is our focus as statisticians, and the demands placed on policy-makers, there is frequently a divide. The divide exists partially due to academic backgrounds, but more importantly because the personality traits that make one successful in policy are generally disjoint from those that make one successful in the hard sciences. Most statisticians are probably not interested in campaigning for public office, and most policymakers are not interested in writing code. Statisticians are the ones who have to close this divide because they are the ones who can. To do this, they need to be continuously moving results, and particularly the presentation of those results, in the direction of balancing technical merit with practical accessibility to policymakers. Too frequently, statisticians jump to their credentials as experts and expect this to be enough to be listened to – and this can mean neglecting to answer the question that's of critical importance to the policymaker. It is addressing that question which shows the policy-maker that statisticians understand their problems. **Schramm** suggested that statisticians need to build rapport with policymakers – learning the nuances of their language, which are different than the nuances of ours, and also learning their dress code. Statisticians should listen as much as they speak, and convince policymakers that they can and do understand their problems. Without trust from the decision makers they support, statisticians will find their technical effort is largely wasted.

John Aston spoke about the statistician's role as a science arbiter. Statisticians have real abilities to understand how bias comes into thinking. And being able to disaggregate when people advocate for their positions is something that statisticians are very good at. **Aston** highlighted the role of statisticians in government; he noted the amazing work that has been done by ONS, but also that there are statisticians in many government departments who have worked tirelessly behind the scenes to make sure that good data was being used in policy decisions. It is important to recognise the many statisticians right across government who, with very little profile, have really tried to make sure that good data is being used, especially given that data was very difficult to get hold of, particularly early in the pandemic.

Some of the government data that has come out of the pandemic has been really useful, but it's been useful partly because it can be complemented by other data that's not government data. The combination between the REACT data and the COVID Infection Survey is a really good example of where bringing lots of different data together has made huge differences to understanding. So, while we absolutely should learn lessons for the future, we should also be very clear that we have done huge amounts of good work and that's only because we as a group of people have come together to support the pandemic response. That is something the RSS should be proud of and emphasise.

5. How did global data feed into the UK's decision-making process?

Sylvia Richardson was part of an interdisciplinary expert group, the International Best Practice Advisory Group (IBPAG), which provided weekly input and challenge to analysis carried out by the International Joint Comparator Unit (ICJU). This unit was established specifically in April 2020 by the Cabinet Office to learn from international responses to the COVID pandemic. Using the, sometimes pretty sparse, quantitative data that were available, as well as information from the diplomatic network of scientific attaches, ICJU produced weekly analyses on a whole range of issues, international trends and PI strategies. These were refined by IBPAG before being presented to Cabinet. We were told that these weekly international analyses were considered as key inputs into UK decision making over the course of the pandemic, particularly when the pandemic was hitting the UK after our comparators.

Nick Jewell pointed to some US resources that could be helpful. A group called OPCAST (the group of science advisors in the Obama administration who informally grouped themselves together again amid the pandemic) issued a series of reports from which the UK can learn lessons. There's also an interesting series of reports from the NASEM Standing Committee on Emerging Infectious Diseases and 21st Century Health Threats.

Andrea Rehman gave a couple of examples of where global data fed into UK border policies. First, the early ascertainment of risks was underestimated from Europe. The Home Affairs Committee (in August 2020) identified a lack of transparency over which data or evidence informed the decision on border management in March 2020. Secondly, the traffic lights system from May 2021: transparent data was published by the JBC and sourced from

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

the WHO as well as a nonprofit, Global Change Data Lab (based in Oxford), who compiled data from government departments worldwide. Initial coordination of global data sources lacked a clear leader: was it ONS? SAGE scientists? The Home Office? DHSC? Or WHO? Evidence submitted to the Home Office preparedness for Covid-19 in July 2021 stated that data sharing between governments was identified as necessary for effective border management, and that there was a big question mark around access to and quality of global data. A question remains as to what has changed in the year since.

The United Nations Economic Commission for Africa commissioned an assessment of the impact of the pandemic on National Statistics Office operations in late 2020 and identified that they were overwhelmed by requests for data and statistical services. These data were feeding into the compiled data sources and forming UK policy. Questions remain over how much support and guidance are available to support global data collection and whether ODA budget cuts indirectly affected data gathering. In terms of lessons, we can learn that transparency is key, government investment in technical sharing of evidence or insight is warranted (especially where data might not always be available), and cuts to international development budgets can impact on data and evidence availability.