

LECTURE

THE ROLE OF AN NSI IN A PANDEMIC

Ian Diamond

Office for National Statistics, Newport, United Kingdom

E-mail: national.statistician@statistics.gov.uk

This article is based on The 2021 Deane-Stone Lecture on Economic Measurement, given on 21 July 2021.

It is an enormous privilege for me to present the Deane–Stone lecture. As a young academic at the University of Southampton, I read so much of the work that came out of the National Institute of Economic and Social Research (NIESR) and was always impressed. Also, as an undergraduate and as a master’s student at the London School of Economics, Sir Richard Stone and Professor Phyllis Deane featured prominently on my reading list. In my opinion, they were pioneers of economic measurement and, today, economists benefit hugely from their legacy.

The focus of my lecture is on the pace of change undertaken by the Office for National Statistics (ONS) over the last 18 months, to enable us to meet the challenges introduced by the pandemic. This rapid change enabled us to fulfil our role as an important national statistics institute—to deliver really strong, trustworthy, accurate data on the economy and indeed on society and the wider world. This is an essential service to all of us as individuals, but also to the service of government, enabling better policy to improve the lives of all citizens.

Of course, although this change in ONS delivery was enacted quickly in response to the COVID-19 pandemic, our methods built upon strong foundations that we had developed in previous years. If we go back historically to the great financial crisis of the late 2000s, the ONS was producing gross domestic product (GDP) on a quarterly basis. Given the length of time between publications, it would be fair to say that the ONS was not able to spot the turns and was criticised for this at that time—rightly so. I would like to acknowledge the great work of Charles Bean in the Bean Report of 2017, which established standards and demands for better and faster economic data. Although it may not be an immediately obvious link, the fact that from 2018 we were able to introduce monthly GDP figures has been very helpful in charting the course of the COVID-19 pandemic. It is also my view that the Bean Report’s recommendation for the building of the Economic Statistics Centre of Excellence (ESCoE), and therefore, the ONS’s ability to partner with the academic community, through NIESR, has been very important. For example, the nowcasting of regional GDP is work that we are very proud to be associated with. Nowcasting enables us to understand what is going on across our nation in a way that is extremely important to policymakers, both at a national and a subnational level and, of course, across the four nations.

At the time of the pandemic’s emergence, we were just developing a new UK Statistics Authority strategy. Sir David Norgrove, the then chair of the Authority, was very keen to entitle it, absolutely rightly in my view, *Statistics for the Public Good*. But underneath that strategy are four pillars: radical, ambitious, inclusive and sustainable, which have been the key elements that have driven our thinking throughout the pandemic. We have approached the questions and conceptual challenges that we faced through these pillars, for example, Do furloughed workers count as unemployed or employed? Do children learning from home count towards output in GDP measures? Is the Coronavirus Job Retention Scheme a subsidy or transfer or government spending? These are some of the conceptual problems that we needed to work through.

At the same time as recognising and responding to what was needed to inform government policy-making, and delivering requirements specific to the pandemic, the ONS was also adapting as an organisation. Like others, we transitioned rapidly to working from home and we adjusted in order to

manage our business as usual. For example, during this period, I am incredibly proud that we were able to introduce double deflation into the Blue Book.

Going back to our response to the pandemic, there were three surveys that we set up extremely quickly which I will describe. People sometimes say that ‘national statistics institutes are wonderful things, but they’re a bit staid and steady’. In this case, I would argue that we absolutely were not. It was very clear at the beginning of the pandemic that we needed to know what was going on in the economy and what was happening to business at a time when the environment was changing very rapidly. So, the first survey I would like to highlight is the ONS Business Insights and Conditions Survey (BICS). This survey, which goes out to a sample of 39,000 businesses and gets a very good response rate, has been published on a 2-weekly basis since early in the pandemic. One example of its utility has been to provide responses on how long businesses think their cash reserves will last. The proportion saying ‘3 months or less’ went up during the autumn of 2020, but, in the summer of 2021, was on the decline. In addition, we are able to subdivide that analysis into different sectors of the economy. For example, in summer 2021, the proportion of businesses with cash reserves of 3 months or fewer was highest in ‘other services’ and lowest in ‘information and communication’. What has been really impressive in a very much one-government way is that we have worked with the Department for Business, Enterprise and Industrial Strategy, HM Treasury and right across government to bring new questions in, and to respond as time went by.

While business, of course, is important, we also need to get the opinions of citizens. So, my second example is our Opinions and Lifestyle Survey (OPN), which we moved to weekly data collection right at the beginning of the pandemic, in March 2020. This is a nationally representative sample of Great Britain (GB) adults and our response rates were over 70 per cent. It is a regular survey—which we close at 4:00 a.m. on a Monday, and then we publish on a Friday morning at 9:30 a.m.—so really pacy data. For example, the percentage of adults who said they were maintaining social distancing was between 80 and 90 per cent in the winter of 2020/2021 but that decreases towards the end of the time period to 62 per cent in July 2021. Again, we worked very closely right across government to address questions that are important for policy and doing so in a way that enables our statistics and analysis to be able to inform that policy. So, for example, we received questions about face coverings at different times, about movement and about loneliness in lockdowns; as well as questions around anxiety for the future—masses of questions which quite simply, you can really only get, I would argue, from a survey of citizens. There is a lot of talk about the role of administrative data, which I will cover later in this lecture, but there are still really important needs in economic measurement and indeed in social measurement for surveys, and the OPN is one that we have used.

The third survey I want to cover is the ONS Coronavirus (COVID-19) Infection Survey (CIS). At the beginning of April 2020, I was at a Scientific Advisory Group for Emergencies (SAGE) meeting where the point was made that we did not know what the proportion of the population who had the COVID-19 virus at any time was, and that it would be unbelievably important for policymaking to know that. So, I took a deep breath and said ‘yes, we will design a survey’. We were able to do this incredibly effectively by working in partnership with the University of Oxford, with Professor Sarah Walker as the joint Principal Investigator, and the University of Manchester, including a brilliant statistician called Thomas House. We also partnered with the Wellcome Trust and the UK Health Security Agency (UKHSA), formerly known as Public Health England (PHE). Working from absolute scratch on a Thursday afternoon, we designed a survey, carefully considered the ethics, planned the fieldwork and went into the field in 1 week. It is a household-based longitudinal survey, taking about 150,000 swabs a fortnight through nose and throat swabs from everybody in a household over the age of 2, and taking blood from a sub-sample of everyone in the household over the age of 8 to measure antibodies.

The CIS has been amazingly helpful in showing the trend over time. For example, in the autumn of 2020, it showed positivity rates climbing, briefly dipping and then the big second peak in January 2021, the subsequent drop down and, in the summer of 2021, increasing again. Having started with a smaller sample size, just focused on England, it soon became clear that these were data that were going to be incredibly important right across the four nations. With this in mind, we had started collecting data for

Northern Ireland, Wales and for Scotland by Autumn 2020 to aid policymaking. But it is not only the positivity rate that was important. Because we were able to study the gene pattern of positive tests, we could see early indications of new variants, that would then be confirmed by the labs. This meant that, by December 2020, we were able to identify within 2 days that we were starting to see the beginnings of a new variant, to be named the Alpha variant, in London and the South East and the East of England. Our data and testing helped us to understand that this strain of COVID-19 was more transmissible and led to the peak observed at the beginning of 2021. The Delta variant has increasingly taken over since mid-2021, and this happened a little bit later in Northern Ireland. What was particularly important was that we were able to understand through this survey, not only the prevalence of COVID-19, but also the variant type of the virus. That has been incredibly helpful.

It has also been the case that we have wanted the surveys to evolve to address emerging and different questions. An example of this is the problems around long COVID. The CIS is a longitudinal household study, which has been very helpful in that we can follow up with people, allowing us to make regular, 4-weekly estimates of long COVID. As of 6 June 2021, 1.5 per cent of the population living in private households had experienced long COVID, which is a serious number of people. We have worked very closely with colleagues and a number of other groups around the country, not only to make estimates of long COVID, but to be able to work to define exactly what we mean by long COVID. Our partners include the University College London, the University of Bristol and the University of Leicester, under Professor Kamlesh Khunti. We have also worked to understand the symptoms, and again, some of the symptoms have changed over time, particularly with those who contracted the Delta variant.

There is an additional point that, while we published data and findings from these three surveys separately on a weekly/2-weekly basis, we have also brought them all together. For example, our CIS estimates of the percentage of adults who have received one or more doses of a vaccine show that this has built up over time. That is fine at one level, but then you need to know population estimates of vaccine hesitancy, which we were able to obtain from our Opinions and Lifestyle Survey. We were also able to subdivide that vaccine hesitancy, for example, by age or by ethnicity, which of course is important in terms of policies and interventions to minimise the hesitancy. In addition, we linked to the views of businesses about vaccination—so in other words, triangulating around different surveys and using the insights to bring together a story around the pandemic.

Another key element to Charles Bean's report in 2017 was that he highlighted that people wanted faster data, particularly with regard to the economy, and I think that is really important. We did start before the pandemic to have a small amount of faster economic data, but we have radically expanded that.

With regard to the economy, faster data on what the consumer is doing is really important. For example, looking over the course of the pandemic, we have used Clearing House Automated Payment System (CHAPS) data, which we bring together with our colleagues at the Bank of England. This data showed that immediately during the first lockdown spending went right down, except for staples. But then people started to spend more on social activities, though this obviously took a long time to pick up, particularly with hospitality being closed. ONS data has been very helpful here. But equally, we have a partnership with a number of credit card and debit card agencies, which has helped us to understand online and in-person spending as well. Understanding what the consumer is doing is an important part of understanding the management of the economy, and I know that our colleagues at the Treasury have found this very helpful.

Equally important is our use of faster retail footfall data from a company called Springboard. We would get this data on a daily basis and produce analysis almost as quickly. Coupled with the timing of announcements about lockdowns or restrictions, these data typically showed the impact of those announcements on footfall. We can then disaggregate it to look specifically at the high street, retail parks and shopping centres. The message during the week of 10 July 2021 was very clear that retail parks had held up better than other places.

A third faster indicator came from the idea to use traffic cameras. These are in operation just about everywhere and there are thousands of traffic cameras that we are able to access data from. Once we have the data, we are able to share it with our colleagues in our ONS Data Science Campus, who are brilliant at tackling numerous kinds of big data-related work. We can then divide these data into cars, trucks, pedestrians and cycles. For example, looking at a graph just for London, what you effectively see is traffic volumes being down for a long time, but picking up from about March–April of 2021 onwards, and we could do the same just about anywhere. This is very interesting in understanding movement and mobility. We have also used telecoms data to do some of this, which again, has been interesting work, but I think the insights from traffic cameras have been very helpful. We have also done work where we link traffic cameras with GPS data on ship movements into ports, to understand trucks leaving ports. That has been useful, not only with regard to economic measurement and with regard to the pandemic, but also with regard to the end of the Brexit transition period.

Another important indicator is vacancy data and here we have partnered with Adzuna, an online job advertisement site, to be able to get real-time data. Comparing 2019 and 2020 data with the incomplete 2021 data, we see a sharp reduction in vacancies in 2020, but a steady increase in labour demand with vacancies going up in 2021. Again, we are able to use these data alongside other data that I will touch on shortly, when we say that vacancies were actually higher in July 2021 than their pre-pandemic level.

Finally, I wanted to mention our use of Google Mobility data. Using searches on Google Mobility, which we could web scrape with their permission, we were able to show people's movements across different settings. For example, compared to a baseline from pre-pandemic times, we saw a real drop in terms of mobility across retail and recreation locations. Parks, however, held up a bit at the start of the pandemic before fluctuating, and there is an increase in residential movements. Following this work, Google now produces mobility trends internationally, and I think it is incredibly helpful.

Indeed, we have been able to develop a whole set of real-time indicators, which can be brought together to inform policy around the economy, to understand how the consumer is reacting. For example, early in the pandemic, we put together an anxiety basket of goods, to enable us to monitor some of the panic buying that was going on. So, for example, in the early weeks, I could have told you that the only place you could get hand sanitiser was in Wick in the north of Scotland. Understanding what is going on in the high street, how hospitality is reacting and how people are responding in terms of travel, for example, to the restrictions that exist—all those kinds of things, and many more, have, I think, reflected our really rapid response to the pandemic. And we are not going to stop doing this. It seems to me that real-time indicators are what people are going to want, in order to understand what is going on in our economy. National statistics institutes such as ourselves should be producing these data and doing so in a way that is accessible and increasingly available for a secondary research to anyone who wants it.

I also wanted to touch on what we have been doing in order to continue to produce reliable, effective official statistics, clearly at a time when we have been challenged in many ways. Not least, for example, if you take our Labour Force Survey that relied on initial contact with our interviewer, our field force going to people's houses and going inside. Clearly, one of the first things we had to do at the beginning of the pandemic was to pull out the field force for the Labour Force Survey; we also had to pull out the field force that went into independent shops in particular, to be able to gauge prices as part of calculating inflation. We therefore really have made enormous use of administrative data. In particular, we have worked very closely with our colleagues at HM Revenue and Customs (HMRC), to be able to use Pay As You Earn real-time information, and to access those data in a way that means we can understand the labour force. Looking at the seasonally adjusted number of employees on payroll right the way through from 2014 you would see it going up steadily, nipping down at the beginning of the pandemic, obviously, and then coming back up. And then you can look at percentage change on the same month in the previous year, across the different regions of England and across Wales, Scotland, and Northern Ireland, and see similar trends. These have been incredibly helpful to us in enabling us to better understand both the demographic and the distributional impacts of the pandemic, and of the restrictions on the labour market, particularly when we have these challenges with the Labour Force Survey on collection and on our response rates.

Focusing in more detail on the Labour Force Survey, what we had to do initially, when we had to come out of the field, was to move to both telephone and online. And in so doing, what we observed pretty quickly was that we had now too many people in a sample who were owner-occupiers. It is a problem, obviously, with moving to telephone interviewing that we basically brought this bias in. What we had to do was bring in a new set of weights based on tenure, and that gave us a new set of estimates, which I think better represented what was going on in the labour force.

I have always believed in open science, and one of the things we have done is to make a whole series of notes available. I think it is incredibly important that we have engaged with colleagues in the academic world to get advice. We have also, if you like, 'put our working down in writing'. And I do think that is vitally important. I would hope that you would trust us that we were doing things that were sensible, and you are more likely to trust us if we are very clear to everybody about what we are doing.

I also wanted to make a point that everything I have covered so far has been at a national or regional level, but we have also really tried to bring the inclusion of individual citizens into our work. For example, when looking at the impact of the pandemic on the labour market through age distributions, using the Pay As You Earn data, you can see that really it is the younger members of the labour market who are the most impacted. There was a drop of over 30 per cent in the number of younger workers on payroll, far, far more than for any other age group. Other work ONS undertook shows the real advantage of being able to link data on inclusion. We were interested in inequalities in COVID mortality, and all we had was death certificates. Now, there is no ethnicity on a death certificate. So, we have linked the death certificate to census data. And that gives ethnicity and also gives us some indicators of disadvantage. We further link it to hospital episode data and to GP data, to understand whether the person who sadly died had some comorbidity. Undertaking a whole series of multivariate analysis, firstly adjusting simply for age distributions, then modelling to adjust for the disadvantage that exists within those communities, shows that you reduce a lot of the excess mortality, but you do not reduce it all. Looking at the first wave going up to September 2020, there was a higher mortality rate remaining amongst black African men, and in the second wave a higher mortality amongst people of Bangladeshi or Pakistani heritage. We believe that this points to further analysis that we need to do, particularly around occupation and other areas, but I think it is important just to show that we are linking data properly to understand the inequalities that exist, in this case in COVID mortality.

I said right at the beginning that I believe that a national statistics institute has a responsibility to all its citizens. We aim to publish our analysis in an accessible way, and we have tried extremely hard to use really good visualisations and good briefings. The bottom line is that we have had enormous numbers of visits to our website, and we do see ourselves now as having a real responsibility to continue to help to improve data literacy across our nation. One of the ways I think we can do that is by continuing to get great data visualisation. Secondly, I want us to be able to use the examples that we have, and ensure they are as relevant and as engaging as possible for people to use. And that teachers right across the educational spectrum can use relevant examples as a way to attract and encourage people to understand data in a better way.

I am not going to go into this at length, but I would like to mention the role of commercial data during the pandemic. Needless to say, it was an interesting challenge measuring inflation when there were not any shops open. Here, we used a decision tree on whether and how to use imputation for items one could normally buy but could not buy now. We were also using web scraping and were undertaking telephone surveys for independent stores. And I think we managed to get some really quite useful data. We also used help from Barclays and from Revolut to weigh tourism data. That helped us to weigh inflation data, and you actually end up with pretty reasonable estimates.

And I finally wanted to say a few words about measuring GDP during a pandemic, where you have got a challenge or two—as I said earlier in this talk what do you do with education? What do you do with health? And so we started to use alternative data sources. So, you are working from home, teachers are working from home, can you calculate, if you like, the full-time equivalent of teachers working from home against the proportion of instruction dependent on the teachers, not on the parents? We used some data that came from an app called TeacherTap, and then we were able to derive a discount for education

that we applied to all remote learners. This discount is less for secondary than it is for primary: it started in April 2020 at about 70 per cent for primary and 45 per cent for secondary; but reduced over time to about 50 and 25 per cent, respectively, by February 2021. I do think we were ahead of many other countries in really trying to understand the impact of the lack of education on GDP. And also, we put a lot of work into understanding health. And you might think, health surely went up, hospitals were full, weren't they? But of course, many elective operations did not happen. So actually, there was a reduction in health delivery, at the time when the pandemic meant that the National Health Service was doing such a wonderful job in protecting us.

So where do we go from here? The first thing people sometimes say to me is that you have been doing an enormous amount of work informing government, does not that impact on your independence? The answer to that is not at all. I see no distinction between relevance and independence. We will continue to be at the centre of policy discussions and informing public debate. To do that, we will talk to government, we will talk to stakeholders and we will talk to the public about what is important. But I can assure you that the methodologies we will use, often working, I hope, with ESCoE, and the statistics that we calculate will be trustworthy and completely independent. And as I say to my colleagues, very often the numbers are the numbers. We will continue to embrace new data sources. Some of the data sources we have used here, I think, can be really exciting. We are going to continue to embrace them, and really have fast economic indicators, while at the same time continuing to deliver the things that, of course, you would want us to, around inflation and GDP and the labour force. But often, maybe using new sources to be able to get better and more accurate estimates. An example of that is some of the work we have been doing recently with HMRC to improve our estimates of migration.

And so to conclude, we will continue to do the surveys, we will continue to use retail information, we will continue to develop sustainable partnerships with people like NIESR and with the public and private sector organisations. We await the results from our Inclusive Data Taskforce, which will require really exciting partnerships with civil society organisations. And we will continue to respond flexibly and at pace, ensuring that we provide data, and sound economic measurement, for the public good.

The above is an edited version of the NIESR Deane–Stone Lecture delivered by Professor Sir Ian Diamond in July 2021. With thanks to Sarah Moore and Bryony Dyssell for their assistance.