



## Public Administration and Constitutional Affairs Committee

### Oral evidence: Transforming the UK's Evidence Base, HC 197

Thursday 9 November 2023

Ordered by the House of Commons to be published on 9 November 2023.

[Watch the meeting](#)

Members present: Mr William Wragg (Chair); Mr David Jones; and Lloyd Russell-Moyle.

Questions 48 to 100

### Witnesses

**I:** Chris Morris, Chief Executive, Full Fact; Hetan Shah, Chief Executive, British Academy; and Gemma Tetlow, Chief Economist, Institute for Government.

Written evidence from witnesses:

– [Full Fact](#)

### Examination of witnesses

Witnesses: Chris Morris, Hetan Shah and Gemma Tetlow.

**Q48 Chair:** Good morning and welcome to the Public Administration and Constitutional Affairs Committee. Today, the Committee is holding its second evidence session in its inquiry into transforming the UK's evidence base.

This inquiry is looking at how officials produce statistics and analysis; how demands for data are changing in the modern, data-driven world; and whether the privacy of citizens has been adequately protected as new and innovative sources of data become available to decision makers. To discuss the perspectives of users of official data and statistics on this topic, we have a distinguished panel this morning. I will ask them to introduce themselves for the record, starting with Mr Shah.

**Hetan Shah:** I am Hetan Shah, and I am chief executive of the British Academy, which is the UK's national academy for humanities and social sciences.

**Chris Morris:** I am Chris Morris, and I am the CEO of Full Fact. We are the UK's independent fact-checking organisation. We try to promote good



## HOUSE OF COMMONS

information in public life, and when we find bad information we try to get it corrected; we try to get mistakes corrected. We are eager users of Government statistics.

**Gemma Tetlow:** I am Gemma Tetlow, chief economist at the Institute for Government. We are a think tank focusing on making Government more effective.

Q49 **Chair:** It isn't quite open questions to begin with. Perhaps we will go in reverse order to the way you introduced yourselves. I want you to give a quick overview of how your organisations use statistics and analysis. Who is it, therefore, who decides what data is collected in the UK?

**Gemma Tetlow:** At IfG, we use quite a variety of data produced by Government. We use some of the ONS data—national statistics—but we also use quite a lot of data produced by Departments on their own activities about themselves. For example, we are interested in statistics on the civil service workforce—the types of skills they have and the types of professions they operate within. We also look at things like the major projects portfolio—what are the big infrastructure projects that Government is undertaking?

On your question of who decides, quite a variety of people are involved in the deciding. If you go from national statistics level, there is clearly a set of important indicators that are decided at quite a high level of Government and some that have been in place in legislation for many years. Something like the retail prices index is in legislation that is decades old. Things like measuring GDP are obviously well established indicators. At the other end of the spectrum on data within Government, it is probably much more ad hoc and lower level. Certain types of data are collected somewhat differently within Government Departments. Something like HR information for Departments is much more a lower-level decision about exactly what gets collected and exactly what the way is.

Also, increasingly, there is a lot of data in the private sector, which we are starting to use more publicly, whether it is things like mobility data from mobile phones, travel data from Oyster cards—it is all sorts of things like that. A lot of that involves decisions made by private companies about what data is collected. It is sometimes not even a decision; the data is just created in the operation of those services. A by-product of that is a lot of data that increasingly could have a variety of uses.

So, on your question, there are lots of people in the system who are making decisions here.

**Chris Morris:** Obviously, evidence is fairly fundamental to what fact checkers try to do. When we see claims made in politics or in public life, we try to seek out the evidence behind them and decide whether that evidence is accurate. If people make claims, we check them; that is across the political spectrum. And in relation to the media, when we do that, we look to a variety of official sources to check the evidence.



## HOUSE OF COMMONS

When we publish fact checks, we are always very careful to link very clearly to those sources of official data, so that people can see where we have got our evidence from to back up the claims we are making when we fact-check people. We always give people a right of reply and ask them for the evidence that they think they are basing the claims they have made on. And if we conclude that we think they have made a mistake or they have got something wrong, we ask them to correct those claims, because we believe that, certainly in public life, accurate evidence is very important.

More broadly, I suppose we see ourselves as part of a broader information and evidence ecosystem. We have, for example, a technology team who have developed a world-leading, automated fact-checking software tool, which is being used around the world. Essentially, it goes through and scopes millions and millions of words every day to find the thousands of words that are important for humans to add context and caveats to. It is starting to be used by fact checkers all around the world, so it is a great British export, I would say. We are quite proud of that.

In terms of how Full Fact sees itself, we want to encourage more political debate, not less, but we believe very firmly that that debate needs to be based on a body of proper evidence that stands up to statistical scrutiny.

In terms of who gathers evidence, I would back up what Gemma says. In particular, evidence from the private sector is, I think, really important. Take, for example, the way big companies gather data. If you are a big company that wants to look at its carbon footprint, for example, you are going to gather an enormous amount of data, both on your internal systems and throughout your supply chains, and that gathering of data can then be very useful to Government—to a Government that is trying to implement the legal requirement to reach net zero by 2050. So I think, when we are talking about the evidence base, it is really important that evidence gathered in the private sector is seen as part of the equation.

**Hetan Shah:** The British Academy is the national academy for humanities and social sciences, so we are a community of researchers; we are also a research funding body. So our researchers are using a lot of Government data. It is worth saying that they are independent producers of evidence as well.

Like my colleagues here, I would encourage the Committee to think broadly about what the evidence system is. It is not just the providers of official statistics. The British Academy itself has done bits of evidence analysis. For example, very early on in the pandemic, we were asked by Sir Patrick Vallance to look at the social consequences of the pandemic, and we produced a major tome on that. That is just one example.

It is worth saying also that I am, with a separate hat on, chair of Our World in Data, which is a website bringing together very long run datasets that are internationally comparable and so we are users of data and think very hard about how to present it.



## HOUSE OF COMMONS

My colleagues have largely covered this question of “Who decides?”, and there are just a couple of things to add to that. One is that I commend the work done in particular by the ONS and the Government Statistical Service during the pandemic and the speed of responsiveness, for example, with the covid infection survey, the dashboards and so on. That shows that these statistics are not just set in stone, as it were; when something happens, we can respond to user need.

I reiterate the point that independent researchers also decide what data can be gathered, as it were. I certainly agree that the private sector is really important, and I hope that we will come back to that. The other part that often gets left out is local authorities. They have very different capabilities. When we are thinking about the UK’s evidence system, we ought to think at that level as well.

**Q50 Chair:** That is helpful. I am certainly going to come back to a number of those issues, particularly the comparability of data. Several submissions from our inquiry have described having access to data as a form of power. Can I ask quickly what you make of that view, Mr Shah?

**Hetan Shah:** There is a lot to say here. I think it links to the question of how data is linked within Government and that accessibility. Independent researchers all say that we can really help to understand what is happening in our country if we have access to that data and if it is linked in a really useful way. Some of that has been improving, but there are still many, many issues. The accessibility is still not always there. We have things like the Secure Research Service, but it is about if we could have more widespread geographical points of access. The API allowing not just individual researchers but actually machine access to that data is still at quite an early stage. The API itself is in beta form. It is working okay. There is not a lot of data behind it, so that would be an area that we would really like to see improve.

If I may, because I am trying to expand the thinking around this, there are some examples of different sorts of data access that would make a real difference. One example is the postcode address file, which this Committee has looked at in one of its previous incarnations. This is now held by the Royal Mail. If this were opened up, it would have huge benefits for UK research and the economy, but it is currently held in one place. That is one example of an issue.

Quite a different issue is archives. For researchers and historians etc, it is about what kinds of archives we have access to. Some Government Departments are certainly quite reluctant to put their stuff into archives.

A third example is the International Energy Agency. We and many other OECD countries are supporters of this; we fund it. Climate change and energy are two of the biggest issues that we face globally, but that data is not currently publicly accessible, so it charges for it. Is there a way of changing that model? Finally, as my colleagues talked about, there is private sector data. It increasingly will be the case that technology companies own this data. How do we use that for public good?



**Chris Morris:** Yes, access to data is a form of power, and so is withholding access to data. It goes without saying, but numbers matter in public life, whether they are about big issues like education or immigration or whether they are about personal health choices and the way that people make those. It is pretty clear that the more data people have access to, the better choices they can make, and therefore the more informed those choices are.

The problem comes when that data is misused or misinterpreted, and then public trust starts to wane. We do not think that there is a crisis in that at the moment. Research done by the National Centre for Social Research in 2021 suggested that public trust in official data is high, but at the same time if you look at things like the Ipsos Issues Index, distrust of Government, politics and politicians, as you know, for the past year or so has been among the top 10 issues of public concern. It is really incumbent on politicians and senior leaders in Government Departments to make sure that data is produced as accurately and, importantly, as accessibly as possible.

**Gemma Tetlow:** On the data sharing, Hetan already mentioned the question of sharing data with researchers outside, and there are some really good examples of that. For things like justice data, the Justice Data Lab brings quite a lot of that together and really helps with research. Education data is also reasonably well advanced in terms of giving access to external researchers, but more could definitely be done.

I do not know if you would call it power or if you could put it down to risk aversion, but there are definitely Departments that do not give as free access to administrative data. Particularly, I have heard of problems where people would ideally like to link data from different Departments.

Q51 **Chair:** Perish the thought that we would ever name and shame Departments, but are there any good—or less good—examples of Departments?

**Gemma Tetlow:** I have just given a couple of examples of Departments that I think are somewhat better: the Justice Data Lab, and the Department for Education's national pupil database and longitudinal education outcomes database. These are pretty well-established research databases that have been used quite extensively.

Other Departments perhaps have further to go. I have heard of occasions, although I don't think I can name them because they are not my personal first-hand experience, where particular problems have arisen from people wanting to link data across Departments. There has been a sense of both Departments being willing to do it, as long as the other Department sends the data and it gets to hold it. That is an area where there is a bit of a problem, or where we need a bit of a cultural shift to accept some of the risks of doing that data sharing.

If I could slightly broaden your question a bit to analysis, as opposed to data per se, we have been concerned about the use of only strategically



sharing analysis within Government, actually as much as outside Government, as an area where that was used as potentially a power strategy.

In particular, we came across this in some work we did, published earlier this year, on the role of the Treasury during the pandemic. As part of that project, we ended up looking at how decisions were made within the centre of Government on some quite important aspects of the pandemic response. It was inherently an issue where there were lots of different issues playing out and interacting with one another, particularly how the economic effects of the pandemic interacted with the public health impacts of the pandemic.

In that report, we raised concerns from the work we were able to do that Departments were sharing analysis strategically with one another in an attempt to put forward their case in opposition to another Department's case, particularly strategic sharing of economic evidence to put forward an economic case from the Treasury versus strategic sharing of health evidence from the Department of Health. We were quite concerned that that was not an effective way of coming to decisions within the centre of Government because of the inherent need to actually properly understand both types of evidence, and in fact to understand how those work together, rather than seeing them in isolation.

**Q52 Chair:** On a similar theme, I am curious to know whether we are counting the things that matter. What are the gaps in the evidence base, and why might they exist? Does anyone have a quick reflection on those gaps? I know in written submissions that Full Fact, for example, has mentioned the lack of information held by Government on its dental workforce—a very topical thing.

**Chris Morris:** Yes, there are gaps, and the gaps mean that there is sometimes a lack of good information about various things in the UK. I think we saw that particularly during the pandemic. Government systems all over the world were clearly incredibly stressed, so it is not surprising, but I think there are some lessons to be learned.

A different example from the one you raised is on something we fact-checked numerous times both from politicians and members of the media. It is the claim that between 100,000 and 140,000 children left school during covid lockdowns and never returned. That is not what the Department for Education statistics actually show; that figure appears to refer to children in 2021 who were out of school more than they were in school. So they may have been in school for 49% of the time, but they were away for quite a bit of it.

The problem is that no one actually knows how to count the children who are missing. The Children's Commissioner for England has said that the Department for Education knows how many and what percentage of children are in school every day, but they do not know if those are the same children every day, if you follow my argument.



## HOUSE OF COMMONS

It comes to the question of whether you are counting the right thing, and whether you are counting something that is useful for people to then make decisions about. Is there a crisis of absenteeism in British schools? Yes. Is the number as bad as has been widely publicised? It is still in *Hansard* and it has been in many media reports. We would argue no. If you are going to do something about it, you should do something about it from a number that stands up to scrutiny.

**Hetan Shah:** There are many, many data gaps, and in some ways I feel sorry for data providers, because our demand for data is now unlimited. If you are the ONS, you have limited resources. I think a top-line point is that if our demands are growing, the resource in ONS and the Government's civil service probably needs to grow commensurately.

To pick out a few, education is one of our big exports, but we do not really have a good handle on what those numbers are; it is only experimental data. Social care has been pointed out as a particularly weak area. With a different hat on, I am on the board of the Legal Education Foundation on justice data, and we have picked out a series of things like court operation, sentencing, remand and bail, where there are gaps.

I would point you back to your own report in 2019—your review of statistics—where you said, “We recommend that UKSA...establish where data gaps persist.” I think it would be useful for you to go back to this—in many other places, your reports have pointed to issues—and to ask what might be done. It strikes me that a few years ago there was the Charlie Bean review of economic statistics, which was wide-ranging. Perhaps there is a space for a wide-ranging review of social statistics, which might comprehensively look at where the data gaps are.

**Gemma Tetlow:** On the positives, I think there have been important new developments in data in recent years. Hetan already mentioned the really quick response during the pandemic to getting things like the covid infection survey up and running, as well as what was originally the business interruption from covid survey, which has morphed over time. There have been some really good examples of where gaps have been identified and there has been a very quick response, and there have been other areas where the ONS, since the Bean review, has been working to develop new types of statistics—whether that is on wellbeing or natural capital.

There are some areas where we are starting to develop the sorts of stats that we have and to reflect new interests, but there definitely are lots of gaps remaining. Chris and Hetan have already mentioned several of them. I would agree with Hetan: from our work, social care is definitely one of the areas where it is very hard to get a handle on what is happening. That is partly because of the way those services are delivered through many private providers, so we just do not have consistent information from there, but we could. We also raise concerns around local government services, where data is collected in a much patchier way and is not totally comparable across areas.



## HOUSE OF COMMONS

To Hetan's point about resources for this, I think the overall answer is yes, we probably do have more data and need some more data possibilities, so more resourcing is probably needed. It is very rarely the case that someone says, "Do we no longer need this bit of data that we have been asking someone to report?" There perhaps could be a bit more of thinking, "Do we need all of what we've got, or do we want to refocus our attention somewhere else?"

**Q53 Mr Jones:** The evidence that we have had so far tends to suggest that the absence of harmonised data, which provides the ability to compare data from various parts of the United Kingdom, is a significant gap in the UK's evidence base. I wonder if you could each indicate your experience of UK-wide data. Do you believe that official sources enable you to compare the experiences of people in the various constituent parts of the United Kingdom? Ms Tetlow, would you like to start?

**Gemma Tetlow:** Sure. The main experience we have of this is in trying to look at the performance of public services across the UK. We do a regular annual report on performance tracking, looking at the performance of key public services, and most of the focus of that is either just on England or on England and Wales because of the level at which those services are managed and the data is collected.

A few years ago, we did attempt to compare the performance of services across England, Wales, Scotland and Northern Ireland. In a sense, devolution since '99 should have provided a really good opportunity to learn from the type of variation that we have had in the way those services have operated across the UK—for example, Scotland has taken a much less market-based approach to NHS provision than we have had in England—but because we don't have comparable data for the four nations, it is very hard to understand what the outcome of those different approaches has been. I would agree generally with the point that it is hard to harmonise data across the four nations, and that makes it harder for us to understand the experience of citizens and to learn from that sort of policy variation.

There have been a couple of positive developments in recent times. The creation of the Office for Local Government should take us some way towards having more harmonised, comparable data, at least within England, for local areas and local services. Hopefully that is a step forward. That office itself will need to grapple with the problem of the lack of comparability in the data, but at least it will be working towards producing a product that means lay users—local authorities themselves—will be better able to know what is comparable and what is not.

Potentially, the private data sources we were talking about may be another way of getting more comparable data across the country, to the extent that those sources are being collected in the same way. Obviously, one thing that creates a lack of comparability between the public data sources, or the administrative data in the four nations, is that because the services are operated differently and because they have different targets





## HOUSE OF COMMONS

for performance and that sort of thing, they are collecting different types of data. That is what drives the lack of comparability.

It would be great if we had more comparable data. There can be a trade-off between the additional administrative burden of collecting data that is not necessary for the operation of the services in different parts of the country. For example, it isn't necessary to collect waiting times in Scotland on the same basis as waiting times in England if Scotland isn't targeting that measure of waiting times, but it would obviously be useful for comparing the experience of citizens if you had those collected on a comparable basis.

**Q54 Mr Jones:** In terms of the citizen's experience, it is essential that a citizen in Wales, where I come from, should be able to assess whether the service that he or she is receiving is as good as another in, say, the north-west of England.

**Gemma Tetlow:** Exactly, but we need to think carefully about the additional administrative burden from asking for data that is not strictly necessary for the operation of the system. Where there is sufficiently high value to understand the experience of citizens, it does make sense to collect it.

**Q55 Mr Jones:** Do you believe that the various Administrations should want that comparability to exist?

**Gemma Tetlow:** Absolutely. As I said, each of the Administrations are running their services in a different way because they believe that is a better way for the system to operate, but they cannot know whether that is the case without comparable data to be able to understand what the practical impact of that is.

**Hetan Shah:** Gemma has hit the nail on the head with what she has said. To add to that, if different services are defined conceptually differently in different devolved nations, it is unsurprising that the data that flows from that becomes more difficult to compare. But there is so much desire to be able to do that comparison for the evidence purposes that we have been discussing. There are some positives. The Government Statistical Service is looking at a statistical coherence programme, so there are initiatives that can help link things up. The ONS, as the biggest of the different bodies, has a leadership role to play. It can reach out to the devolved nations and try to set that standard.

The other thing worth flagging is that international comparability is not just about the devolved nations; there is a much bigger playing field. There are still questions about our relationship with Eurostat post Brexit, because it would be helpful to make sure that we have that comparative data with our European counterparts. Coming back to the research base, there is a lot of interesting evidential research happening right the way around the world. One thing that the British Academy focuses on is our nation's language-learning ability, because we are starting to get to the stage where we are not able to read the stuff being published in the



## HOUSE OF COMMONS

journals to know what other countries are doing about their educational health systems.

- Q56 **Mr Jones:** You indicated that the ONS is doing some good work, but do you think that, by and large, comparing the experiences of citizens in various parts of the UK is becoming more or less challenging 20-odd years after devolution was implemented?

**Hetan Shah:** Yes, I think so.

- Q57 **Mr Jones:** Sorry, I mean do you think it is more challenging?

**Hetan Shah:** It is certainly becoming more challenging. That is not at all a surprise because if countries diverge in what they are doing, that's what I would have expected to have happened. You therefore have to do more to try to keep that comparability.

- Q58 **Mr Jones:** In the middle of all this, there is the poor citizen who is unable to assess whether he or she is getting as good a service as someone in London—in my case—or in Edinburgh. Mr Morris, do you want to add anything?

**Chris Morris:** I endorse what both my colleagues have said. You inevitably have to conclude that the fragmentation of the production of statistics across the UK is not an aid to public understanding. From our perspective, that is a problem. I also think that if the way statistics are produced is fragmented, that inevitably leads to misunderstanding and mistakes, because they are produced in different ways.

- Q59 **Mr Jones:** It must be a particular problem for you in your particular job, I would have thought.

**Chris Morris:** It is, because it is very difficult. For example, it is far from ideal that you cannot reliably or easily compare mortality rates across the UK—in Scotland, they are collated slightly differently.

Again, when it comes back to the issue of public understanding, many people who might come across these things on occasion, rather than as a professional necessity, probably do not notice whether a statistic refers to England and Wales or to the whole of the UK. That leads to misunderstanding and to data being produced that is not being consumed in the way that the producer probably hoped it would be. Hetan mentioned some things that the GSS is doing, but it would be good if the UK Statistics Authority could be given more of a mandate to ensure that there is much more comparability of statistics across the UK.

- Q60 **Mr Jones:** That would have to be by agreement among the four Administrations.

**Chris Morris:** It would, indeed.

- Q61 **Mr Jones:** Do you suspect there is a reluctance to come to that sort of agreement because it might cause embarrassment to certain Administrations?

**Chris Morris:** Devolved government is our system of government, so that is the position we are in. You may have a better idea than me about whether it might cause embarrassment, but it should be in everyone's interests, from our perspective, that the reliability of information and the production of statistics is as accurate as possible for what people believe is their country, whether that be one of the four nations or the UK as a whole.

Q62 **Mr Jones:** Do your individual organisations make representations for more harmonisation of data? Is that something that you do?

**Chris Morris:** We have discussed it with the statistics authority on occasion, rather than with devolved Government.

Q63 **Mr Jones:** Because it seems to me that that would be a useful service you could all perform, frankly. It is frustrating your work, isn't it?

**Chris Morris:** Yes, noted.

Q64 **Mr Jones:** Ms Tetlow, in recent months, the ONS has revised key GDP estimates and has flagged issues with employment numbers. Should we be concerned about the quality of the UK's economic data?

**Gemma Tetlow:** I think we should always be concerned about the quality of the economic data, given that it is a really important part of understanding what is going on in the country, and policy clearly responds to the data. On the particular issue that occurred recently with GDP data, as Ian Diamond outlined to you in his evidence, the particular challenge was in the periods when the economy contracted very sharply—the start of the pandemic—and then expanded very rapidly when restrictions were lifted. Those are unusual circumstances for trying to get a handle on data.

Over recent years, we have moved to a situation where the initial estimates of GDP are produced more quickly, the desire there being to have as quick a grasp as possible on what is going on in the economy. That is good in principle, but it does mean you are trying to produce those statistics on the basis of less data than you ultimately have available to you. That seems to have been the problem in this case—some of the information just was not available in those initial estimates, and that became apparent later on. That was exacerbated by the particularly unusual circumstances of that contraction and then expansion.

There are probably two things that we ought to look at going forward. One is whether aspects of the data that became available later on could be made available more quickly. That could be by sharing them, for example, from HMRC to the ONS more quickly, or by HMRC getting that information more quickly from businesses to the extent that it relates to tax returns being filed. There would obviously be some trade-offs, in terms of the cost and what business would be asked to do.

The other approach would be to say, "Do we think the early data are too unreliable and that they would actually cause more harm than good?" I don't feel in a position to judge that. However, we have clearly tried to move in a direction of having at least some data available as soon as



possible, in order to help politicians to make decisions or to help the country to understand what is going on, rather than having no information. I think that is probably where we want to be, but there is a trade-off and we need to make sure that people understand the uncertainty that surrounds those initial estimates.

**Q65 Mr Jones:** On balance, do you think that the ONS should have done something differently?

**Gemma Tetlow:** I don't feel well enough informed. I can't say that I was sitting here when those first estimates came out saying, "It's clearly the case that they don't have enough information to produce an accurate estimate", and I don't remember anyone else saying that at the time either. However, ex post we have clearly learned that there were challenges for those particular periods with the data that was available.

How much we can learn from that about what to do in the future, I'm not sure. I guess that the question is how much we think we learned from that situation that could be carried over to future experiences, or whether it was very specific to the behaviours that were going on around those closing-down and opening-up periods.

**Q66 Mr Jones:** Generally speaking, do you consider the UK's evidence base to be trustworthy and of high quality, Mr Shah?

**Hetan Shah:** Yes. I would say that a lot of the statistics that are produced are of high quality and are trustworthy. We have a badging system that allows these statistics to be called national statistics and many of them achieve that quality. We have the Office for Statistics Regulation checking them, regrading them and so on, so there is that kind of check and balance.

There are certainly areas where there are worries. One worry that all statisticians are facing—this is not a UK problem; it is around the world—is falling response rates to surveys. With the Labour Force Survey, for example, over the course of a decade we have seen the response rate fall from 50% to 15%. We're not alone in that regard, but collectively we need to think how we are going to tackle that.

There are some particular issues that could be ironed out. For example, there are some areas where one Department's data does not cohere with what another Department's data suggests, so if we could just get those two Departments to bash their heads together and come up with what is actually going on. On self-employment, the numbers from DWP and HMRC suggest slightly different things.

On earnings, there are discrepancies between the DWP and the annual survey of hours and earnings, or ASHE. And there are some areas, like business data or trade data, that academics in the UK will tend not to work on, because the data quality is not high enough and therefore they can't get published in good journals. It feels to me that if you want the analysis, you need to have the good data underneath it.



## HOUSE OF COMMONS

One thing to flag is that this isn't just about data; we also need to think about the data infrastructure in the UK. There are some old systems—legacy systems—that now need upgrading, and some out-of-date software.

The other thing that worries me is that every company in the country now wants to hire a data scientist and most statisticians are rebranding themselves as data scientists, so how will the Government hold on to talented people? That is an issue for the future.

**Chris Morris:** On the specific issue that you raised about GDP revision, it seems to me that the problem there is one of communication rather than one of errors in the initial data, which were incomplete. Clearly, the size of the revision was politically unsatisfactory, but it seems to me that it is incumbent on the producers of statistics to make clear the caveat that this is very early data, but it is also incumbent on the media and the markets not to just jump on the headline figure without explaining the context.

Q67 **Mr Jones:** Yes, but in the absence of any caveat, it's not unreasonable for them to jump on that data, is it?

**Chris Morris:** Well, the caveats have to be there and that would be my more general point. I think that there is extremely good-quality data in this country. I think that the issues, certainly from our perspective, are in communication, presentation and explanation of all those contexts and caveats.

There is another comparison that I think is worth making. We work with a lot of fact-checkers elsewhere in the world and we shouldn't underestimate how lucky we are to have a reliable system of statistics that is politically independent. In so many countries now, the statistics have become so politicised that it's actually quite difficult to find an independent body of evidence on which you can check facts. It is something we perhaps take for granted, and clearly it is not perfect and improvements can always be made, but we are lucky to have a system that is politically independent. I think it would be unfortunate if the ONS essentially became a target for political criticism.

Q68 **Mr Jones:** You really put your neck on the block when you challenge people on data, do you not?

**Chris Morris:** Well, we have to get it right. It is kind of implicit in the name Full Fact, is it not?

Q69 **Mr Jones:** But you are relying on official statistics. In the case of this particular example, I do not know whether you relied upon this data in challenging anybody, but it could rebound on you quite embarrassingly, could it not?

**Chris Morris:** It certainly could, and if we make mistakes, which we obviously do, we are very clear to say so publicly and explain what the mistake is. It is obviously worth saying also that it is not just official statistics that we rely on—we rely on a whole body of other evidence—but, yes, we do rely on official statistics to be accurate. As I say, compared



with some of our colleagues elsewhere in the world, in countries that would be considered democracies, I think the reliability we have in front of us is something we should not take for granted.

**Gemma Tetlow:** A further thought on that—I think Chris’s point about communication is right—is that the ONS does try to communicate some of the uncertainty around statistics. It will present standard errors and the like. They are perhaps not terribly front and centre in the release for obvious reasons, and there is a bit of a trade-off with communicating to the public in a way that is understandable. However, on the GDP numbers, for example, there was a lot of focus in the media coverage about whether the UK economy was back to its pre-pandemic size and was it or was it not the only country in the G7 that had not regained its pre-pandemic size?

In truth, given the uncertainty around the estimate, there was some probability essentially that the latest numbers suggested that we were back above our pre-pandemic size. I did not see anyone cover it that way in the media, so to Chris’s point, I guess it would be helpful if people who are then conveying that information, rather than putting a lot of weight on the narrative that the UK is the only one not back at its pre-pandemic size, had instead conveyed more of the uncertainty around that, saying, “There is a probability we are not, and that comparable probability for other countries is x.”

Q70 **Mr Jones:** But in order to do that, the producer of the data would have to issue the caveat we have just discussed. As I understand it, it did not on this particular occasion.

**Gemma Tetlow:** I believe the ONS stats will provide the standard errors around its estimates alongside the data, so you could work it out from that, but it obviously requires a bit more interpretation.

Q71 **Mr Jones:** It was not so explicit though, was it?

**Gemma Tetlow:** No. I would have to double-check, but I do not think the ONS was the one highlighting “not back to pre-pandemic levels” in its releases.

Q72 **Chair:** No one has mentioned Disraeli’s quote so far, so we are doing very well in that respect. I just want to ask something going from what Mr Shah said, and this is directly to Mr Shah: how effectively do officials work with researchers and academics outside of the Government to ensure that everybody involved in this is improving it in a joined-up manner? What is that relationship like?

**Hetan Shah:** I would say that it is still too weak and needs improving. I would say that the Government Statistical Service and the ONS have better linked up with their Government users. The Office for Statistics Regulation in its most recent “State of the Statistical System” report said that “engagement with a wider breadth of users of statistics is still too limited”, and academics were a particular area it pointed out.

It feels to me that this is an area where the Government should be doing much more outreach. I wonder whether the UK Statistics Authority as a



board could lead on that, because individual statistical departments can do that, but if we saw the UK Statistics Authority itself at board level reaching out and doing a big exercise to say, "Who are our users? Can we come and talk to you?", that would go a long way in signalling all that.

- Q73 **Chair:** Building on that—this is perhaps for Mr Morris and Ms Tetlow—from your experience professionally, how do statisticians and analysts across Government work with organisations like yours to understand what data is needed and therefore what should be collected?

**Gemma Tetlow:** I think the practice is varied. My experience is that some parts of Government have better links to academia and research than others, and so therefore they have stronger links to getting insight into what evidence is needed.

- Q74 **Chair:** Do you think, for example, that there is a robust framework in place that aids officials in making that decision about what is useful, valuable or sought-after statistical evidence to seek?

**Gemma Tetlow:** That is a tricky question. The code of practice for statistics is reasonably clear about the quality and considerations required in producing official statistics. I think that the problem for me comes around it not being clear what should or should not be considered an official statistic. There are an awful lot of types of data and broader analysis that Government can do—sometimes does do—where I do not think there is terribly clear guidance on what should be made public or how decisions are made about the relative priorities of different things.

The ONS has been putting extra effort into certain types of data and in filling certain types of data gaps. For example, there is new work ongoing to try to improve measurements of public service productivity, which is clearly a current Government priority, with John Glen leading the initiative on that. It is not clear to me how Government makes decisions about resourcing something like that versus some of the other data gaps that there might be, or how external asks for new data are balanced against internal governmental priorities.

Statistics is one area in which, actually, a lot of the benefit is long term rather than short term, so we need to be conscious about ensuring that resources do not get too skewed by very short-term Government priorities. Often the benefit of that data collecting and improving measurement is much more long term rather than very short term.

- Q75 **Chair:** I'm sorry to interject another question at this juncture; please feel free to come back to the one already asked, but just building on what you have said, there appears to be a perception in some quarters that Government views often drown out other voices in the conversation about what data should be collected. You mentioned short-termism there; is that a concern that you would recognise?

**Gemma Tetlow:** As I say, it is not—I am not sure that I could concretely point to something that has been clearly an inappropriate use of resource compared to other demands, but it is not at all clear to me how that

prioritisation happens, either across the whole of Government, with ONS resources or even how Departments think about the relative priorities on their time.

**Chris Morris:** On the first point, obviously every system can be improved, but our experience is that the system in place is reasonably robust and a lot of it is encapsulated in the code of practice for statistics. The problem is that you can have a reasonably robust system on paper, but it is about how that system is then implemented.

For example, the ministerial code says that Ministers should bear in mind what is said in the code of practice for statistics; we would recommend that that language should be changed to say that they should adhere to the standards contained in that code of practice, because there is no point in having a code of practice if everyone ignores it.

On the second point, about whether Government voices drown out others, I am not sure that I would use the term “drown out”; I think it—

**Chair:** It was an emotive term, sorry.

**Chris Morris:** Yes, well, I think that it is right and proper—and inevitable, in fact—that the Government should be the loudest voice in the room in all its forms, but I think that some thought needs to be given to who else should be heard. If you can get voices from the voluntary sector, the private sector and so forth, you will be looking at the same problem but from an angle that Government may not think of, and the more perspectives that you can get, the more you are going to cut bias and inequality out of the system and the better the evidence that you will collect.

Q76 **Chair:** Thank you very much. That is very helpful. Mr Shah?

**Hetan Shah:** Just to give a specific example, I was on the cross-party Social Metrics Commission, which Baroness Stroud chaired. That has come up with a new measure of poverty that has been widely accepted by civil society and so on, but actually getting it going within Government has proven really quite slow and difficult. I think that that is a very good example of this. We have done a lot of work externally but, because we are not within the system, it is much harder to get that picked up.

Q77 **Chair:** I am just hypothesising here, but is the reason that poverty is such a politically charged issue? Do you think it comes down to the politics of it?

**Hetan Shah:** It is always a mixture of things across different issues, and inertia is often as much to blame. People are very busy and so on, and finding the space to do yet another new thing when there is already a lot on your plate can be as much of an issue.

Q78 **Lloyd Russell-Moyle:** Chris, what does it mean to communicate data well? We have heard a bit about some of the failures of the ONS and others to communicate it with all the nuance. What would you say is





doing it well?

**Chris Morris:** From our perspective, it needs to start with accuracy above all, and then communication needs to mean that the data you produce is accessible in different forms. Perhaps Gemma is very happy when the data is in reams of spreadsheets.

That is important because researchers have to have access to that raw data, but it is also important for producers of statistics to think about how to present that information to a much more general audience. That means, when it is in written form, a lack of jargon and using clear language. There also needs to be more thought put into things like data visualisation. We now live in a visual culture; that is how people learn and consume information.

An example is when the Foreign Office put out a data graphic talking about the amount of Russian bank money that had been sanctioned at the beginning of the Ukraine war. It was a big number—£258 billion, I believe. When we looked into it, it was not at all clear where that number had come from and how it had been calculated.

To be fair to the Foreign Office, when we talked to them about it, they revised the guidance they put out alongside that number. One of the problems with data visualisation is that it is attractive, but it is even harder to get the caveat and the context in because you are literally just putting a few numbers on a page, so you have to think even harder about whether you are using the right number, because it is not a written submission in which you can add that context and caveat.

Data visualisation is really important in communicating well. Looking further ahead, I would actually go a lot further. Why should there not be some sort of virtual ONS that people can enter and interact with as though it were a video game? I am not suggesting that a virtual ONS would be as popular as “Minecraft” or “Call of Duty”—I think that may be pushing the ONS a little far. I suppose my point is that you cannot have analogue communication in a digital world.

Q79 **Lloyd Russell-Moyle:** What do you mean by a virtual ONS?

**Chris Morris:** As in a video game, you can go in, be in a virtual library and pick out the statistics you want. That is how my children’s generation obtain their information.

Q80 **Lloyd Russell-Moyle:** Is that not a bit like how the census data has been presented, with those maps that you can zoom in and out of? Do you mean more of that, or is there something else?

**Chris Morris:** Yes, more of that, but using the next generation of technological advances, which are almost virtual reality data. Yes, absolutely, that will bring data and information on which people can make informed choices to a much broader section of the population. Alongside the raw data in spreadsheets and so forth, a lot more thought needs to go into attractive and quite groundbreaking presentation. It is difficult stuff to



do—I am not saying anyone has got it right—but the Government certainly has the resources to begin to make a difference.

- Q81 **Lloyd Russell-Moyle:** Hetan, do you think the presentation of data is improving or getting worse? New forms of communication make things more instant, and sometimes harder. As Chris just said, sometimes visualisation can make things less clear.

**Hetan Shah:** I would say on balance that it is getting better. There are more opportunities to interact with a wide range of users, including directly with the public, and statisticians should be commended for having leapt into this new world.

There are more blogs, social media posts and so on, but I think you could go further. The ONS website, which you pointed to in previous reports, could still benefit from better navigation design and engineering. You could be publishing user guides on what data exists in each field of statistics, where to find it and so on. You could even have compendia for social trends or economic trends. These things used to exist some years ago. The difficulty is almost that there is so much stuff now, so how do you make sense of it?

- Q82 **Lloyd Russell-Moyle:** You need someone curating it, but who takes the responsibility to be fair and impartial? Is it the ONS or someone else? Who then funds that?

**Hetan Shah:** I think the ONS or statistical bodies have a role to play in at least saying, “What are the guardrails of what is interpretable?” We as a democracy can then argue and think-tanks can say, “This is good. This is bad” etc, but at least telling us what is within the bounds of agreed facts would be really helpful.

The other point I want to make on the question of communication is that sometimes we look just at the ONS. I think that when Government talk about policy, they should also be publishing their chain of reasoning and evidence underneath that. Sometimes we focus just on the data piece, but actually looking at more evidence transparency across policymaking is a real opportunity. There is an opportunity for the Committee to think if this is a potential inquiry around evidence-informed policy and what could be done around that.

- Q83 **Lloyd Russell-Moyle:** Do you think that on the whole, policymakers and public figures use statistics, data and numbers well, or is there a lack of understanding by some of the individuals in those posts?

**Hetan Shah:** There are several things to say here. One is that probably since the creation of the UK Statistics Authority and the ability of the chair of that body to write to people and say, “You have used these badly” etc, that has led to an improvement across the piece. I think there are fewer misleading statistics out there, and for the most part people are a bit more careful. I do not think that we should expect every politician to have brilliant statistical skills; it is more about taking care and ensuring that



they can show that there is a chain of reasoning. Where does that data that they have talked about come from, and so on?

I talked about evidence-based policy. I think the other area is around improving evaluation. Again, I feel like that is something that the National Audit Office has pointed to: lots of projects are done, but actually 8% of Government spend on major projects has had robust evaluation. It is not just about individuals; it is about systems.

- Q84 **Lloyd Russell-Moyle:** You mentioned the letters from the UK Statistics Authority. Are they a kind of stick threatening to beat any politician or public figure that uses it incorrectly? Are there things we could do beyond that that would encourage better use and understanding of numbers?

**Chris Morris:** In answer to your initial question, obviously there are good examples, but quite often numbers are not used well. One of the problems is the reluctance to admit when mistakes are made. I think it is incumbent on Government Ministers and Departments to correct their mistakes when they are made and do it very publicly, because that generates further public trust. That is one point that is important. It is also important in Government Departments for the tone to be set politically by Ministers and, within the Departments, by permanent secretaries to make sure that staff are sufficiently trained and understand the importance of what they call intelligent transparency. It is really important, when numbers are not used well, that that is made very clear. That is an important way to retain public trust.

- Q85 **Lloyd Russell-Moyle:** There is a danger, though, that people will use those corrections themselves as signs of political weakness—

**Chris Morris:** So we need to have a political system that looks at correction as a strength, not a weakness.

- Q86 **Lloyd Russell-Moyle:** We have just had a quasi-set of questions around, "Did the ONS fail in the fact that it amended its statistics?". Rather than us saying, "Isn't it great that the ONS came forward and amended that?", we are all saying, "Well, the ONS must have messed up in the first place." If that body was headed by a Minister, it would have been even worse, because a political claim could have been made that these were treated as something other than just technical processes. Does having letters that are public—a stick—sometimes hinder the ability of someone to then be honest and say, "Actually, maybe there was another way of doing this", without it becoming a personal attack?

**Chris Morris:** Our argument would be no. We are critical of MPs, in particular when we ask them to correct things and they do not. What I think is important is that Full Fact would like to increase trust in politics, not diminish it. To do that, politicians need to take the lead, because with public service comes public expectation and public responsibility. By being honest when mistakes are made, you create trust with the public. In particular, that applies to numbers and the numbers that people use in their political communications. I don't think people should be afraid to say,



"We got something wrong." I think it is a sign of strength, not of weakness.

- Q87 Lloyd Russell-Moyle:** Gemma, Chris talked about this virtual reality world that we could start to do, and Hetan has talked about the need for curators to understand how people can sift between the good and the bad to work out what is data. How do we ensure that people have the skills or are supported to understand what the underlying data means, rather than just what is presented to them on a nice little meme?

**Gemma Tetlow:** Perhaps I can first pick up on a couple of the points discussed earlier—particularly the point that Hetan made about the transparency of the evidence base used to inform policy; I agree that that is particularly lacking at the moment.

When Ian Diamond gave evidence to you, he talked about the concordat to support research integrity and his aspiration that Government would be compliant with that. But from our work, it seems that Government analysis is currently a long way from being compliant with that. There is very little transparency around a lot of the analysis that is done within Government to support policymaking, which, as I think Chris said—maybe it was Hetan—often makes it difficult to understand the chain of logic from the evidence that Ministers are looking at to the decisions they then make.

To your point about the ability to accept that there was a mistake and to learn as new information comes along, I agree with you; I should have said that when we were talking about the ONS revisions. I think it should be seen as a positive that when new information comes along, we can update and make improvements. You could imagine a different world, in which Government was much more transparent about the analysis that had underpinned decisions. We could open that up to the outside world, which would allow much more peer review from experts outside Government.

- Q88 Lloyd Russell-Moyle:** This is an assumption that decisions are being made through a rational, data-driven approach, rather than a political, ideological approach, which sometimes confounds data and rationality.

**Gemma Tetlow:** Being an analyst myself, I would hope that policy decisions are informed by evidence. You can distinguish between the political objectives that you might have but none the less be informed by solid, data-based analysis of, "These things are going to achieve this particular political objective." I think you can distinguish that from simply wishful thinking that this policy may, in some way, achieve the political objective that you had, with absolutely no evidence base to back that up. I am not sure that is a good role for politicians to be in.

- Q89 Lloyd Russell-Moyle:** We have seen all the WhatsApps from the covid inquiry about some of that. Are you suggesting that when policy statements or Bills are produced, alongside them—like we have the explanatory notes—there should be a set of statistical notes to show the evidence bases that we have used? If you are suggesting that, what kind of political narrative do you put alongside it? Or do you expect it just to



## HOUSE OF COMMONS

be the statistics, with no political narrative at all?

**Gemma Tetlow:** Yes, essentially that is what I am talking about. I worked with Sense about Science some years ago on an evidence transparency framework, trying to look across different Departments at the extent to which it was possible to see the evidence that had informed different decisions. I am not sure it needs a political narrative; I think the evidence base should support the decision that was made, although I guess you could add more narrative around how that evidence was interpreted to reach the political decision. But it is more about helping the outside world to understand the evidence base that was used.

As I say, there are probably at least two potential benefits from that. One is helping the public to understand why the decision was made, and Sense about Science pointed to some particular problems that arose during the pandemic. For example, covid guidance was coming out but, for people on the frontline, sometimes that conflicted, so, without being able to know what the motivation for the guidance was, it was hard for them then to make a decision about how they interpreted that in their own particular circumstance on the frontline.

Also, as I said, for peer review purposes it would be beneficial to Government itself for it to open up its analysis to experts outside, who could say, "You know what? Actually, there is a different method you could have used" or "This bit of data would give you a better answer to this question," and then you could use that to inform updates.

Q90 **Lloyd Russell-Moyle:** I suspect that, during covid, part of it was just that something needed to be seen to be being done and it wasn't based on any data, but that is how a lot of things happen. We have had numerous statements, including even in the House today, about things that are like that. I want to go back a bit: how do we ensure that civil servants, Parliament and beyond have the skills to be able to navigate that?

If we are publishing this or even just dealing with the data that is already published and out there, how do we make sure that people—particularly those involved in decision making—have data literacy? I ask because it's quite easy to present a set of data that can look like it demonstrates your point if you don't know how to dig deeper into it.

**Gemma Tetlow:** This is something that we have looked at in our work both on the civil service and on Ministers, and—I think this is what you're getting at—for policy-making civil servants and for Ministers, it is not necessarily the case that they need to be deep data experts themselves, but they need to have enough skill to be intelligent consumers of that data and to be able to ask the right questions.

Our work on the civil service suggests that at the moment the approach taken across different Departments is quite patchy and not all that effective in some places. So some work has been done to do a bit of a—it has been described as a "sheep dip" of one day's data training for policy professionals. Our view is that that is probably not a terribly effective way

of doing it and actually it would be better to shift the resources on to something more targeted.

Also, there is a bit of a lack of an overarching strategy within the civil service about what data skills are required and where those data skills exist. This goes back to the question earlier about gaps. We do not have good data on who in the civil service has those skills, so when it comes to a new problem, it is sometimes hard to identify who has the skills to work in the team to interpret the data most effectively. So there are things that could be improved there.

For Ministers, there seem to be, again, disparate approaches across different ministerial offices. In some recent work on private offices for Ministers, we found some examples where Ministers are starting to have data scientists in their private office team, to have a very direct resource for interpreting the data, but other Ministers don't have that.

One thing that we found a bit more concerning was, in some of our work with Ministers, finding that they don't always know who in their Department can provide that sort of data analytical support to them. They don't always know that they have a chief scientist, chief economist or chief data officer. Perhaps you could do more with the induction process for Ministers, just so that they know who the people who can help them to interpret the information are.

**Q91 Lloyd Russell-Moyle:** Some of those people need to be based and located next to Ministers for them to be called upon.

**Gemma Tetlow:** Potentially, although some of our work from relocation of the civil service suggests that you can have effectively people disparately located as long as you have those connections and people know who—

**Q92 Lloyd Russell-Moyle:** Let's see. I am more sceptical about that, I think. Being physically next to someone is much better than being down the line.

I want to move on to something slightly different, which is the census. The ONS proposal is to end the 10-year census and instead produce more frequent estimates using alternative sources. What do you make of that? I would like to hear the views of all of you as to what you make of that. In particular, are there any concerns in terms of data that you would have about moving to that model?

You mentioned earlier some of the difficulty with administrative data. Is this a problem or is it a positive thing, and are there particular areas that we would need to be aware of? We'll start with Gemma if that's okay.

**Gemma Tetlow:** I think there are definitely positives. The major positive is having more up-to-date population data than we currently have from the census, which is only every 10 years, and then sort of interpolated between. Potentially, it could also save you significant amounts of money from doing a decennial census as well.



## HOUSE OF COMMONS

I think some of the potential issues with administrative data need to be taken seriously. We would need to have a bit of a shift in the frequency and openness to sharing administrative data for these purposes, because it would clearly be very crucial that the ONS had access to all of those in a timely manner.

I guess we have not touched so much on the fact that because of the way administrative data is collected, it may not always capture, for example, all the people who you would need to see. The ONS is having to think very carefully about who might be missed from this approach, who you get from the census.

The other area of concern would be those aspects of data that are harder or not captured at all in administrative data sources: for example, occupation, sexuality, religion, and languages spoken, which I don't think are regularly collected in any other form of administrative data source. We need to be careful that if we lose something, we are happy to lose it, and that it is a trade-off we are willing to make. I think there is potential. There are good things about this, particularly as we have had a more dynamic population with people moving more and higher levels of inward and outward migration, but there are challenges in a different approach.

**Q93 Lloyd Russell-Moyle:** Is there a nuanced approach you could take to this, where you keep some sort of door-by-door survey or census, but in a much more reduced format?

**Gemma Tetlow:** I am afraid I don't know enough about how that would work.

**Q94 Lloyd Russell-Moyle:** Chris, what are your views about it?

**Chris Morris:** I don't have a huge amount to add to what Gemma said, in terms of the detail of the ONS proposals. Anything that makes information more up to date will obviously be welcome for people like us. We do use the census a lot in our fact checking. It is an important resource, including the ability to knock down false claims about things like the ethnic make-up of our big cities. It is an important resource for us. The main point about the census where it does quite well, as you have already mentioned, is the accessibility of the information and the way it is displayed. In terms of the proposals, it is probably not something that I have particularly up-to-date information on.

**Q95 Lloyd Russell-Moyle:** If religion, opinion, sexuality and maybe even racial make-up are the things that we are saying will be much harder to assess through administrative information, and you are saying that those are the bits that you probably rely on most heavily with the census, then there is a gap there that we need to ensure is filled before we go headstrong into this.

**Chris Morris:** As Gemma mentioned, the thing to work out first is where the gaps are. We have already talked about data gaps, which is absolutely critical. There are always going to be gaps. How do you count things and are you counting the right things? Those are the questions we need to ask.



## HOUSE OF COMMONS

**Lloyd Russell-Moyle:** And who is doing the counting.

**Chris Morris:** Yes, and who is doing the counting.

**Hetan Shah:** The ONS has just closed a big consultation on this, and it will be very interesting to hear what they have got back. I agree with my colleagues that this is something we must be exploring in comparative countries like New Zealand, Australia and Canada, all thinking about this exactly for the reasons that have been said about timeliness of data, cost and so on.

It is worth recognising that the 2021 census was at a very weird time, right? It was during the pandemic, so that benchmark data we have about how people were travelling to work and so on is quite skewed. We are now going to sit on that for 10 years. You can see why this is something important for us to be thinking about. There have definitely been strides made in using administrative data and matching that data, although we do not have a population register, which is I suppose what statisticians here would love to have. It would make their lives much easier, but we're not going to have that.

As we have slightly explored earlier today, admin data is reliant on the underlying policy. Take, for example, child benefit. Back in 2011, when the ONS was starting to think about all this, it probably thought that child benefit was an area where it would be able to rely on that as a way of picking up data. Actually, the changes to policy in 2013 meant that that probably was not such a good dataset for them anymore. How do you protect against policy change that impacts your fundamental data? That is one question.

I think Gemma's point about the most marginalised missing from admin data is really important. These are the people who are the least likely to interact with your health service, education service and so on, so how do you track for that? What is interesting at the moment is that the admin data we have is relatively consistent with national and regional pictures at census level; where it breaks down is at local level. That is what you really need to safeguard, because there is strong user demand for that local-level information. I think that will be really critical.

Gemma pointed to identity and protected characteristics. As one example, data on religion and religious background is required for fair employment monitoring in Northern Ireland. How would we close that? We need to find those gaps.

I suppose if I were you, from where I sit there may be four tests to think about. One is that this cannot just be a cost-saving measure: it must actually be for statistics. Secondly, the admin data needs to be strong enough. That needs to be proven and take into account the local area, capturing the most marginalised in society and so on. Thirdly, the admin data must be sufficiently stable. Are there some binding commitments that need to be made, be that legislatively or not, that would actually safeguard those datasets? I gave child benefit as one example.





## HOUSE OF COMMONS

Finally, from my own backyard, there is finding some way for record preservation for historians. If you look at what historians do with the census, it is incredibly important. It would be a real shame if you moved admin data and we did not have those datasets for the future.

- Q96 **Lloyd Russell-Moyle:** So we need to do some of that underpinning here in this place, because at the moment the census is the only data source that is set and agreed by Parliament. Is that something that will need to be agreed?

**Hetan Shah** *indicated assent.*

**Lloyd Russell-Moyle:** Thank you very much for that.

**Chair:** Thank you. We have a quick question from David Jones.

- Q97 **Mr Jones:** Isn't the weakness of the census and, indeed, all survey data that the responses are, of necessity, subjective? Ms Tetlow, you mentioned the question of "What language do you speak?". As you will probably know, there is a distinct variation between the ONS's number of Welsh speakers and that of the Welsh Government. It depends on what the individual respondent considers to be speaking the language. You could speak fluent Welsh, you could speak restaurant Welsh—if there is such a thing—or you could just be able to say "good morning" or "good afternoon" in Welsh. Subjectively, even if you can say only a few words, you may consider yourself to be a Welsh speaker. Is that not a weakness of such data?

**Gemma Tetlow:** Yes, it certainly is. I think the census also asks, "What is the primary language that you speak at home?", which may be a more useful indicator in understanding the proficiency of people and what language they are most comfortable communicating in. For some of these things, I am not sure if you could collect the information in anything other than a subjective way unless you are actually going to require a language test for everyone in the country, which may be—

- Q98 **Mr Jones:** So you would have to apply a gloss to the information that is provided to take account of the individual respondent's own perception of his or her ability to speak the language.

**Gemma Tetlow:** I think it is always going to end up depending on what you want to use the information for. With something like Welsh language proficiency, you could want to use that to demonstrate the widespread speaking of the Welsh language. That would be one question, in which case you would want to interpret the answers differently.

You may want to understand how many citizens there are for whom Welsh is the only language in which they are proficient, and therefore, if you are delivering public services, do you need them to be communicated in Welsh as well? That is quite a different question, and you would want to interpret the answers to those kinds of questions very differently. For example, do they also report speaking English?

- Q99 **Mr Jones:** Very few, I would have thought.



## HOUSE OF COMMONS

**Hetan Shah:** If I might add to that, I think this is where statisticians really come into their own, because they are very aware of the limits of the data they gather. For the census, they test the questions pretty well, especially the ones that we have had for a longer period of time. We will then benchmark that kind of data with other sources, as it were, so it is exactly that: if there is a discrepancy, they think it through.

**Chair:** Thank you. We have one very quick, final question.

Q100 **Lloyd Russell-Moyle:** You have talked about compatibility and making other data sources compatible with these administrative data sources. When you go to your appointment, the NHS records, “What is your sexuality? What is your ethnicity?”, but we do not have a centralised way of recording that. Is that maybe what needs to happen—for some of the subjective filling-in by people to be collected somewhere other than their house?

**Gemma Tetlow:** This is exactly what the beyond the census programme is trying to work out. As Hetan said, the challenge is that we do not have a population register; we do not all have an ID number. Some of the challenges here are about ensuring that we link the same person from the NHS records into their tax or benefit records and dealing with the fact that not everyone will interact with every service. If you have not been to the NHS, do we then not observe some of that information about you?

**Chair:** May I thank our three panellists today for their extremely useful evidence and for sharing their expertise? If there is anything further that you wish to let us know about, please do write to us. In the meantime, thank you very much for your time.