

Public Administration and Constitutional Affairs Committee

Oral evidence: Transforming the UK's Evidence Base
, HC 1682

Tuesday 5 September 2023

Ordered by the House of Commons to be published on 5 September 2023

[Watch the meeting](#)

Members present: Mr William Wragg (Chair); Jo Gideon; Mr David Jones; Damien Moore; Tom Randall; Lloyd Russell-Moyle; Karin Smyth; John Stevenson.

Questions 1 - 47

Witness

I: Professor Sir Ian Diamond, National Statistician.

Written evidence from witnesses:

- [UK Statistics Authority, Office for National Statistics, and Office for Statistics Regulation](#)

Examination of witness

Witness: Professor Sir Ian Diamond.

Chair: Good morning and welcome to the Public Administration and Constitutional Affairs Committee. Today the Committee is holding its first evidence session in our inquiry into transforming the UK's evidence base. This timely inquiry will look at how officials produce statistics and analysis, how demands for data are changing in the modern world and whether the privacy of citizens is being adequately protected as new and innovative sources of data become available to decision makers.

This morning we are joined by the National Statistician, Sir Ian Diamond. I wonder, Sir Ian, whether you would introduce yourself for the record, please.



Professor Sir Ian Diamond: Good morning, Chair. I am Ian Diamond. I am the UK National Statistician.

Q1 **Chair:** Thank you very much. Before we come on to the things I outlined in my introduction, as we have the opportunity of having you before us, could I ask about the recent revision of the GDP figures? What is your thinking as to that? I just really want some context, if you would not mind explaining it to us, please.

Professor Sir Ian Diamond: I would be delighted. International best practice is that we get GDP figures as quickly as we can, but we recognise that that is not based on the complete data. Central banks and Governments like an initial view very quickly. We then produce subsequently what we call the Blue Book, which is the revised figures based on the very best data.

The figures we put out last week were for 2020 and 2021. I do not need to tell the Committee that this was a period of great turmoil in the country, as we went through the pandemic. That means the data on which we have based our revisions have been subject to a lot of scrutiny by ourselves and a lot of checking, but also some of the data have come later.

As I said, this is international best practice. Other countries will be producing revisions over the next little while. We are producing our revisions perhaps a little earlier than some other countries.

I also do not need to tell you that there are eight quarters across the period 2020 and 2021. We did not revise six of those quarters. The two that were subject to significant revision were quarter 2 2020—that is April, May and June, which you may recall was the height of that first wave of the dreadful pandemic—and quarter 2 2021, as we started to recover really quite quickly.

I might then move on and talk a little bit about what the new data showed us and explain a little bit about where that came from. There were four main areas that caused the revisions that we brought out last week. First, we had underestimated the extent to which firms were stockpiling. We just did not have the data. Firms really did have a lot of reserves.

Secondly, the costs facing firms in some areas—flour, for example—were going up, and overall the costs facing firms went up.

Thirdly, both retail and wholesale margins held up better than in our early estimates. There may now be some really interesting work to do to understand that. I stress that what I am about to say is speculation, but that could well be to do with some of the interventions that were made by the Government to protect the economy during the period.

The final one was the extent to which our wonderful NHS bounced back and did rather more operations or outputs, as we would call them, than



HOUSE OF COMMONS

we had initially expected. If you put all those together, at a time of enormous turmoil, yes, we have made a revision, and it is a positive one.

I also have to say that some of our initial work was rather different to some of the work in our sister NSIs in other European countries. We tried to measure the extent to which education and health had been impacted by the pandemic. Other countries just assumed that input equals output, and therefore the amount of money being spent was the same as the output. We tried to measure the extent to which education was impacted, which meant our initial declines were a little bigger than some other countries and so now we have seen a return.

Q2 Chair: The variation is 1.8%, which is significant. It is a very politically important matter as to whether this country was the “sick man of Europe” in terms of the recovery or was firing on more cylinders than we thought. You are very mindful of the political significance of this, or of how your statistics are used by politicians.

Professor Sir Ian Diamond: I spent August completely concerned with this, as have my colleagues. What we call our curiosity sessions have been in overdrive. We really wanted to be sure about what we were saying. We knew we were going to be under deep scrutiny about this, but we wanted to be sure we would not have to revise. We are very comfortable with the numbers we have.

Q3 Chair: Perhaps particularly in terms of comparable European systems or economies, are you aware that those revisions are going on? Do you have any indication that their GDP figures are going to be revised upwards or indeed downwards?

Professor Sir Ian Diamond: No. Given all the issues of pre-release access, William, I would not tell you the numbers ahead of their publication. I can assure you I have not told the French or the Germans in advance either and nor, sadly, are they ringing me up.

As soon as any of them publish, we will make sure we give you a message to let you know, “This is a new one”.

Chair: You will forgive me for my curiosity, I am sure.

Professor Sir Ian Diamond: I am equally curious—do not worry—but I know what the answer would be if I made a phone call.

Chair: Yes, that is understood.

Q4 Mr Jones: Sir Ian, is it fair to say that this episode has been rather embarrassing for the ONS?

Professor Sir Ian Diamond: I would not say it was embarrassing. I would have to say, as I started my remarks, that this is international best practice. We would expect to revise. We cannot revise on the basis of no data, and some data take some time to come in. The alternative would



HOUSE OF COMMONS

have been not to have produced data. That is a worse place to be. Central banks and Governments would not want us to do that.

Certainly, perhaps we might spend more time on the day explaining this to anyone who wishes to listen, but sometimes people just look at the headlines.

Q5 Mr Jones: There has been adverse comment in the press, as you probably know. The *Spectator* commented yesterday, "Indeed, this blunder makes you wonder: if they can get this one wrong, just what else are they cooking up or nodding through?"

Professor Sir Ian Diamond: I have said this on a number of occasions. It is very easy to say, "If they can get this one wrong". We have followed international best practice to produce the best estimates that are possible to give with the data that are available at a particular time because central banks and Governments would like to get those answers reasonably quickly.

There are data that are simply not available. When they come out, you then revise. That is international best practice. It is not a mistake. It is a revision. I recognise that you may think I am playing with semantics in doing that, but, if you had asked me two years ago or when we originally published those data, "Will these data be revised?" I would have said, "Of course, yes. I am expecting them to be". That is what we would do.

I return to the point I made. Of eight quarters at a very difficult time, six of them we did not need to revise. Particularly where the big decline was and where the big recovery was, we have had to make some revisions because the data simply were not there.

I am sure you read the article by Tim Leunig in the *Financial Times* yesterday. He said the ONS data are fine. He said the ONS statisticians are fine. Interestingly, the main thrust of his argument, which is a reasonable one—we have already been thinking about it—is about whether we can get the data that we have to wait a long time for any quicker. I waited until the team that had been working flat out to get these numbers out had finished. My first port of call in September was to say, "Let us get together and see how we can get some of these data quicker". Some that would require data sharing, and I hope we are going to talk about that at some stage later this morning.

Chair: We are indeed.

Professor Sir Ian Diamond: That is going to require quick data sharing across Government in some areas. If that could be done, we might be able to get some of those data quicker. That would mean the revisions would potentially be a little less because we would have some of the more accurate data earlier.

Q6 Chair: Thank you very much indeed for that. Going on with our agenda as planned, could you briefly outline which areas of Government are



HOUSE OF COMMONS

responsible for producing statistics and analysis and which parts of those systems you have responsibility for?

Professor Sir Ian Diamond: Statistics are produced right across Government in Departments and indeed in some cases in arm's length bodies. Those statistics are subject to a code of practice.

I have responsibility in two different ways. I have formal responsibility for the Office for National Statistics, which produces a large number of data. We have just been talking about some of the economic data that we produce. I then have dotted-line responsibility, through my leadership of the Government Statistical Service, for all the other statistics that are produced.

In each Department, there will be a head of profession for statistics, who is responsible for producing those statistics. She or he is part of the heads of profession for statistics in the Government Statistical Service. While their line management responsibility goes vertically through their Department, I have a pastoral responsibility and a responsibility to make sure standards are met across the Government Statistical Service.

In the rest of the UK, again, I have very strong dotted-line relationships with the chief statisticians of Scotland, Wales and Northern Ireland, but, again, statistics in those Administrations is a devolved responsibility.

Q7 **Chair:** What is your involvement in the drafting of the national data strategy? Do you have any role in the ongoing delivery of that strategy?

Professor Sir Ian Diamond: We were involved, as in ONS was involved, in contributing to it. It was being drafted formally within DCMS. We offered our views and we were part of it. I have a responsibility for mission 3, which looks at data standards and data sharing. I drive that forward, but I do not have the overall responsibility for the national data strategy.

Q8 **Chair:** How do you work with the Cabinet Office's Central Digital and Data Office?

Professor Sir Ian Diamond: We work extremely positively. Megan Lee Devlin, who is the head, and I meet on a regular basis. It is the central place where the frameworks and some of the standards are put together. While it is a fragmented system, it is a system that we work hard to make sure is coherent.

Q9 **Jo Gideon:** Sir Ian, we have received evidence describing the UK's data landscape as "highly and unnecessarily fragmented". Is that true?

Professor Sir Ian Diamond: It is certainly fragmented. If you think of the way I have just described it, there are different statistics decisions being made in all four Administrations of the United Kingdom. If I then may move to Westminster, each Department has its own statistics.



HOUSE OF COMMONS

Having said that, the Government Statistical Service does work very hard to make sure we are producing statistics to the same standards. The heads of profession, if they require some advice or support, are very good at ringing me up or having meetings with me. Together with the chair of the UK Statistics Authority, Sir Robert Chote, I have spent a lot of time visiting each Department to engage with statisticians, to listen to them about their challenges and to give advice.

You could do things in a slightly different way. In Canada, for example, statistics is, in its terms, a federal responsibility. We have a system that we have to work very hard to make work, and I will continue to do so. Some of the other questions that are likely to come will highlight some of the challenges we face. For example, statistics are produced through the code of practice, and that is great, but I have no influence on which statistics are produced, for example.

Q10 Jo Gideon: Can you please tell us about the rationale behind the establishment of the Government Analysis Function?

Professor Sir Ian Diamond: Yes, the analysis function was established before I joined this role. It was set up at a time when Government were really looking at a matrix model. You still have the Departments but there is much greater strength. I am a huge fan of this model in this case. Instead of each Department having its own function and doing it in its own way, we try to get standards that go right across Government. I do not need to tell you about the Government Finance Function or the Government Commercial Function.

The proposal that was made, which was a good one, was to say, "Let us bring together the professions who are doing analysis on a regular basis: actuaries, economists, operational researchers, statisticians and social researchers. Let us bring them together and say that this group can have similar standards, similar career opportunities and bring better quality to Government across the piece". That was a very good point.

Subsequent to that, the geographers asked to join. That was agreed. Just yesterday we held a town hall for data science to discuss where data science was, what data scientists would like to be and whether potentially—and I use the word "potentially" deliberately—they might want to join the analysis function.

You may say, "The statisticians are different to the economists", or whatever. It is worth saying that in a Department, on the ground, analysing data, you will often find an operational researcher, an economist and a statistician working on the same project, sitting next to each other and driving things through. It does make sense.

At the same time, we have to recognise the unique properties of each profession. If I take those professions I have just described to you, there is an intersection in the Venn diagram between them. There is no question. There is quite a big intersection in the Venn diagram between



HOUSE OF COMMONS

some of them. At the same time, there are distinct features of each one that mean it is right to be a profession.

The aim of the analysis function was to bring standards, to bring career advice, to work together, to make sure the whole was bigger than the sum of the parts and to make sure that, where there were particular skills that were needed, the function could work on them and bring them to the table.

Are we there yet? We have really moved on. When I took over this role, I would say the vision was there, but the journey had not yet really started. We have moved quite a long way in the last couple of years. We are now coherently together. In many ways, you might say, "Hold it. How can you bring these together?" If you think of the Government Finance Function, there are disparate professions within the Government Finance Function and, again, they work together as one.

That is the vision of the analysis function: Government get better quality analysis across the piece and we are able to have some sensible conversations with, for example, the policy profession. My own view, which is shared by Tamara Finkelstein, who is the head of the policy profession across Government, is that the very best policy is not made by analysts sitting in darkened rooms, thinking up some analysis and metaphorically throwing it over and hoping a policy professional catches it. It is made by teams of policy professionals and analysts working together to identify the questions that need answers and the data that are needed to answer those questions and developing, in an iterative way, better analysis, which leads to better policy.

I have been privileged in my career to do that on a number of occasions. I am very clear that analysts benefit enormously from spending time working with the policy professionals and the potential beneficiaries. It makes better analysis. That is what we are trying to achieve. I am not going to pretend to you that the journey is over—is the journey ever over?—but I have to say we are really making good progress.

Q11 Jo Gideon: I was going to ask you whether the analysis function has achieved what you hoped it might.

Professor Sir Ian Diamond: It has not achieved what I hoped it might. It will achieve what I hope it might. We are making good progress. My view is always that you need a crisp enunciation, which I have tried to give you: improved policy, better analysis right across Government and higher standards. You need to know what success looks like and you need to know broadly what the route map is that you are on to get there. Are we there? No, not yet. Have we made good progress? Yes.

Q12 Jo Gideon: Was the establishment of the function a sign that the distinction between statistics and other forms of analysis is unhelpful?

Professor Sir Ian Diamond: I do not think so. It is over 50 years since I took my first undergraduate course in statistics. During that time, I



HOUSE OF COMMONS

have worked with and engaged with people from many different disciplines.

In some of the work I did in the 1980s, I was developing things that the economists were developing at the same time. Microeconomists were developing very similar models. They were just calling them different names, essentially, and maybe using slightly different methods to estimate them. It is not new. This is my point about the intersection in the Venn diagram. A lot of microeconomists use a lot of very good statistics. I have worked with operational researchers who are using the same or very similar methods. It is not an unhelpful distinction.

It is incredibly helpful that different professions with different ways of thinking—operational researchers think in a different way—and different training are approaching the same problem in slightly different ways. You end up with better answers because you can toss things around and really improve the analysis.

It is not an unhelpful distinction. There is a distinction between being a statistician, being an economist and being an operational researcher, but it also benefits everybody to interact and work together.

Chair: We have a set of quite existential questions now from David Jones.

Professor Sir Ian Diamond: I would have expected nothing less.

Q13 **Mr Jones:** Sir Ian, could you tell us how analysts decide what is properly categorised as a statistic and therefore subject to the standards set and enforced by the OSR? What are some different species of analysis that may be used to inform Ministers, for example, but may not necessarily be made available for public consumption? Do these analysts always get these decisions right?

Professor Sir Ian Diamond: That is existential. Those are really important questions. It is absolutely right that, when you do analysis, you might work on a whole set of different issues. Some of those may not subsequently become policy, but you want to be able to try things out. You are answering a “what if”, and you might have five different “what ifs”. You then are able to present an options analysis, and one becomes policy. It is entirely reasonable that those kinds of analyses take place.

When something becomes a statistic is when you are saying, “We have met all of the standards of the OSR and we now have a national statistic. We are going to put that into the public domain and explain exactly how we came to it”.

Your question is, “Can you get that wrong?” It is hard to know. There is a subtle difference in asking, “Can you always get it right?” There is a standard question about when we put things into the public domain. That is a reasonable one. If it is going to be used, if a Minister or a politician is



HOUSE OF COMMONS

going to talk about it, or indeed if I am going to talk about it, it needs to be of the right standard and in the public domain.

We do sometimes use the term “experimental statistics” when we are developing things. I will be honest with you: that is not a term I particularly like because “experimental” suggests we are not quite sure they are any good. Actually, they are very good, but we want to make sure we can get them absolutely right before they become national statistics.

In summary, analysis needs to be going on all the time. When you actually want to talk about it, it then has to be in the public domain. I could give you an example. For example, during the pandemic, we were producing estimates of positivity. If I was not happy, as happened just two or three times over two years, with the quality of the data that were coming from the laboratories—maybe the laboratory had changed its machine or something—we would say, “Hold up. We are not happy”. If I did not publish on the right day, I had to ring up the regulator and make that very clear.

It is very good that we have a system that says, “We will announce when we are going to publish. As soon as it is spoken about, it needs to be in the public domain, at the same time as everybody else has it”. At the same time, there is a lot of analysis going on day in, day out, which is exploring and trying to understand exactly what is going on in a system.

Q14 Mr Jones: The Committee has seen a number of examples of pieces of analysis being produced by officials that are not then being made publicly available or perhaps being published after considerable delay, apparently not really taking into account the needs of users. Would you agree, therefore, that the standards of the analysis function are weaker than the code of practice for statistics?

Professor Sir Ian Diamond: I am not sure I completely agree with that, but I have always said very clearly that things should be published. I am very clear in my mind on that. I have always said that we should, as Government, be part of what we would call the Concordat to Support Research Integrity. Patrick Vallance, when he was Chief Scientific Adviser, and I signed that. That says that we will publish the results of the work we have done.

For example, I believe very strongly that, if you are going to do an evaluation of a policy, you should announce your protocol and put that in the public domain; you should announce when you expect to publish; and you should publish it. That should be the norm. Having those data not only published but available for secondary analysis is absolutely essential.

Q15 Mr Jones: Would you accept that, as I said, analysis is apparently produced and not published, or may be published some considerable time later?



Professor Sir Ian Diamond: I would accept that it is probable that some analysis is not published, although I do not have details of that right across Government. I would accept that some things do take some time to get out. My view is that, when the work is started, the publication date should be made available so we know when to expect it.

Q16 **Mr Jones:** Is there a case for strengthening the analysis standard?

Professor Sir Ian Diamond: Following this conversation, I will go back and have a very careful look at it and respond to you. Initially, it is probably not far off the right place. Let us have a careful look at it.

Mr Jones: Perhaps when writing to us you could give an indication as to how you think that could be achieved.

Professor Sir Ian Diamond: Yes.

Q17 **Tom Randall:** Sir Ian, I wonder whether I can ask you about emerging sources of data. As I understand it, statisticians have traditionally used surveys to collect data and more recently have moved to using administrative data, and then more recently still new forms of publicly owned administrative data are being used, like geospatial data and online scrape data and so forth. I just wondered whether you could provide examples of where these newer sources have made real changes to the evidence base and also to decision makers.

Professor Sir Ian Diamond: That is an absolutely important point because we have moved into a world where the whole question of what our data is is a really good one. The other type of data you did not mention, if I may say so, is textual data, which we now have the ability to analyse in a way that is better than it has been in the past.

We need to use administrative data and born-digital data widely, but let us just remember a couple of points, and then I will come to some examples, if I may. Administrative data are collected for administrative purposes. When I design a survey, I worry greatly about whether I am asking the question in the right way. I worry greatly about the question that is being asked and whether or not people will respond to it appropriately.

Administrative data are collected for administrative purposes, not to answer research questions. They can be incredibly helpful. Often they can be faster, for example. I always say very clearly that you need to look terribly carefully at exactly what those data contain. If you believe that administrative data are collected without error or bias, I am afraid you are perhaps on a different planet to the one I am on.

I will give you an example. We use really regularly—they are great data—HMRC data on numbers of people on payrolls. We can get them very quickly, on a weekly basis, and they give a fantastic indication of what is going on in the labour force.



HOUSE OF COMMONS

By their very name, they do not include everybody. Those data are the people on payrolls, so they do not include the self-employed. For example, if there are policies going on that might encourage people to become self-employed or to return to the labour force, you may not be seeing exactly what is going on in the labour force. It may be biased in that way. Would I therefore not use them? No, they are incredibly important. We get them much quicker and they are very helpful indicators, but you need to understand exactly what the issues are. Those are really good data.

You mentioned geospatial data. There is enormous potential to use telephony data, anonymised completely, to understand movement. During the lying in state of her late Majesty the Queen, telephony data were used to monitor the speed of the queue and also to identify when to close that queue. That was an operational policy. Those data were set up and done very quickly.

Let me explain one of the things we are working on. We have a new agreement on telephony data. When I talk to local authorities and I talk to people here in Westminster, they will say, "It is really super-interesting and super-important to know the number of people who are usually resident in Westminster", which I can give them, "but what we want to know is how many people are in Westminster at midday today". Of course, that includes people who have commuted for work, people who are visiting for tourism or whatever. We believe that using telephony data can give us much better estimates of commuting and the number of people. That becomes incredibly helpful to local authorities in planning services, and we will do that.

We have a major project going on at the moment using scanner data from seven major supermarkets. These are billions of pieces of data, but you might imagine that, if we use them, we can get better, faster and more granular estimates of inflation. We are not there yet, though.

We have just brought on in February much better estimates of train prices—it is a real advance—using electronic data. Over the next couple of years, we will be bringing on board scanner data to replace the traditional methods of measuring inflation, which use a person with a clipboard in the supermarket. You may say, "How on earth did you manage to do that during the pandemic?" We did not. We had to take that out and we substituted it using web scraping. You rightly mentioned web scraping.

Those kinds of administrative data and born-digital data, web scraping or using innovative sources, can be incredibly helpful and can often produce data quicker. They are also sometimes much more inclusive because they include everybody. That can be incredibly helpful. At the same time, they are subjected to bias and error. You need to be very clear about what has been collected.

Q18 **Tom Randall:** Just building on that, you say the data is useful but it is



HOUSE OF COMMONS

subject to error and bias. When you are doing a survey, you know what you are collecting and the state of play when you are collecting it. Does using these other sources of data introduce new ethical concerns?

Professor Sir Ian Diamond: Yes, absolutely.

Q19 **Tom Randall:** How do you manage that?

Professor Sir Ian Diamond: Interestingly, on bias, I was working with colleagues in Scotland last week around the final figures for the Scottish census. I do not know whether you are aware, but the Scots will announce their census results on the 14th of this month. It is a really superb piece of work to have done. The whole conversation there was about bias.

To move to the ethical concerns, typically, when you do a census, you ask for consent. The person has given you consent. Certainly, when you are doing a survey, you have consent because the person has answered the questions. You will often ask for consent potentially to link them to other sources. If you are doing a cohort study, you might want to link them to the previous cohort or to link them to other sources. You can ask.

When you are using administrative data or indeed born-digital data, very often there is an implicit consent. That is why it is incredibly important that the use of any form of data such as the ones I have described, which have the potential to have incredibly positive impacts on all our fellow citizens, is regulated. It has to be regulated. The way we do that is by asking, "Do you have ethical permission?" For example, there is a National Statistician's Data Ethics Advisory Committee, which looks at all the ethical issues around doing no harm and whatever. There are many very good ethical committees across our country.

Secondly, we ask, "Is it in the public interest?" The strap line of the UK Statistics Authority strategy is "statistics for the public good". I believe it is very important that any normal person can see what is being done and that any normal person would think that it is in the public interest that that is done.

Thirdly—this comes back, if I may say so, David, to your really important question about transparency—we have to publish this stuff. It has to be transparent. What we are doing should not be hidden somewhere. You need to know this work is being done using these data.

Fourthly, as I said at the start, we use these data with the implicit consent of the public. We should not assume that. Clearly, every time you want to use data that have been collected nationally, you cannot go and ask everybody, but you need to have a very strong process of public engagement. I believe you also need public involvement in the decision making.



HOUSE OF COMMONS

We do have to take all these issues into account. The use of radical new datasets and the possibility to link some of those datasets offer immense opportunities to answer questions that we have only been able to dream about answering and to answer questions that have the potential to impact positively on the lives of our fellow citizens.

We should only do so if those data are analysed and put together ethically, if privacy is maintained, if we are transparent in the way we use them, if it is absolutely in the public good and if there has been a programme of public engagement around it.

Q20 Mr Jones: You have mentioned the issue of devolution already today. Given that administrative data are drawn from operational systems that frequently vary among the four administrative parts of the country, will the increasing use of these sources in our official statistics exacerbate issues around the comparability of data?

Professor Sir Ian Diamond: It is pretty difficult to say they will exacerbate the issues, to be honest. It is not very easy to make comparisons already.

Q21 Mr Jones: With more use of it, presumably the problem will get worse.

Professor Sir Ian Diamond: The problem may stay pretty bad, to be honest. That is a disappointment, personally.

During the pandemic we had a really good example in the coronavirus infection survey of a survey that was designed UK-wide. It was designed to enable comparisons. It was very helpful to understand some of the things that were going on in different parts of the UK to make policy. We also run what we call an opinions and lifestyle survey, which again is designed to be UK-wide and can therefore be used to make comparisons.

You are entirely right that statistics is a devolved responsibility and therefore the data that are collected for administrative purposes in different parts of the United Kingdom differ. We have found it very difficult recently to collect comparable data for different administrations across the UK on the health service, for example.

Q22 Mr Jones: I was going to ask you about that. The Secretary of State for Health, as you probably know, recently wrote to the Welsh and Scottish Health Ministers drawing attention to this problem. He said, "I also believe we need to work together to ensure that health data is more comparable across the UK. It is important that all our citizens can understand the performance of the health services they are receiving and that we can learn from what has been tried and tested in one part of the UK to improve services across the country".

He then said, "I welcome the work our respective teams have been doing to improve data comparability, for example through the Office for National Statistics' work to improve key UK-wide health performance metrics".



HOUSE OF COMMONS

These are very important issues. I am a Welsh MP, and I am finding it very hard to understand how well the health service in my part of the country is performing in comparison, for example, with the health service just down the road in north-west England.

Professor Sir Ian Diamond: I completely concur with that observation. I am so pleased the Secretary of State for Health wrote that. We have been working very hard to try to get comparable data. Comparable data are possible in some areas but not in others. Trying to get cancer outcomes is very difficult because they are collected in different ways.

It is an opinion that is made there that it would be good to be able to make those comparisons. I recognise the opinion that you have. Many people say to me that it would be fantastic to be able to evaluate the national experiment that is devolution by having data that are directly comparable. We can do that. While statistics is devolved, I do not have the ability to ensure that all data are collected in a way that is comparable. We work really hard to make comparable data as best as possible, but at the moment I have to be honest that not all data can be compared. For example, I cannot do ambulance waiting times.

Mr Jones: You mentioned that in Canada this is a federal competence.

Professor Sir Ian Diamond: That is right.

Q23 **Mr Jones:** In retrospect, was it a mistake to devolve the competence for statistics to the devolved Administrations?

Professor Sir Ian Diamond: It is not for me to comment on that one. If you as politicians want those data to be comparable, one way would be to give me the power, as in Canada, to produce statistics that are comparable.

Q24 **Mr Jones:** Is it fair to say that the current arrangements make your life harder?

Professor Sir Ian Diamond: They make my life harder if what you want is comparable data.

Mr Jones: It is a reasonable ambition, surely.

Professor Sir Ian Diamond: I accept it is a reasonable ambition. If you are asking for comparable data, I find it difficult to say I am not able to get there because of the way in which they are collected in different parts of the United Kingdom.

Q25 **Karin Smyth:** Just to pick up on that, I have looked at health data before. It is not just the statistics being devolved. The organisation of those services is different as a result of devolution. Going back to our earlier conversation, data that is collected in an administrative way will be different to looking at the outputs. You would have to not have devolution of health in order to be able to make that comparison.



Professor Sir Ian Diamond: There are some issues there because, of course, if you are making a comparison in the first place, you should be comparing like with like. If I may move to education, the Scottish education system is very different to that in England, for example. What has been done after three years is not comparable to what has been done after three years in a different Administration, clearly.

I am not for a minute suggesting that the solution to all comparable data is simply to say, "We will make statistics a UK-wide issue". I would never say that. It is incredibly important that we are in a position where we are able to compare apples with apples.

Q26 **Karin Smyth:** Following on in that vein—we touched on this earlier when we looked at your involvement in how the statistics are produced—we are interested in looking at what can be done to improve the value of administrative data for statisticians and analysts. Should statisticians be involved in developing the systems from which you require data and how would that work?

Professor Sir Ian Diamond: There are a number of questions there, on most of which implicitly I agree with you. When I say "implicitly", I just want to make sure I am agreeing with you.

Let us start with this question. If you are going to collect a new set of administrative data, should you have statisticians involved? Yes, 100%, for the very simple reason that one of the things statisticians are very good at is framing questions.

There are statisticians whose profession is question design. It is incredibly important because sometimes you think, "That is a really good question", but people answer it in a way that is not the way you might expect. If you ask people, "Do you have a job at the moment?" many self-employed people will answer, "No, I do not have a job. I am self-employed". You have to ask in a slightly different way. It is incredibly important to involve statisticians and to involve the people who do behavioural tests on questions in the development of administrative systems. That is the first point.

The second important point is that, if we want to get more from those systems that exist, we might want to look very carefully at the data that are collected. There is a really strong case at the moment for a review of what is collected in administrative data to ask the question about whether we can get all the things we would like to get.

If we are thinking about moving away from a census, for example, it is really important that we are able to use administrative data to get the information that users would want. That may require making some changes to administrative systems. In my view, that requires the involvement of statisticians right at the beginning because they are the professionals who know how to answer questions.



HOUSE OF COMMONS

It has always been one of those difficult questions throughout my life. Asking questions is something we do day in, day out. People who are absolutely not statisticians believe this is pretty straightforward stuff and anyone can design a quick questionnaire.

In my life, I have had the challenge of trying to make sense out of data. People have come to see me and said, "I have designed a questionnaire. I have some data. I want to answer this question using these data". You have to say, "I am really super-sorry. This questionnaire does not answer that question". You then either have to say, "What could we do?" or sometimes, sadly, "There is not a lot we can do here".

Involving the professionals in the design of questions, both in administrative systems and in the surveys that we do all the time, is incredibly important.

Q27 Karin Smyth: Taking a practical example, you talked about the census. Can you briefly summarise the proposals you have recently made regarding the provision of data around population and migration in England and Wales? What sort of feedback have you had to date?

Professor Sir Ian Diamond: We are in the middle, as you are aware, of one of the most exciting consultations we have ever done.

I do not know whether anybody has been to Corfe Castle. I was in Corfe Castle recently. My family were having nothing to do with this, by the way. They went off to a café, and I took William Morton Pitt's 1790 enumeration and went around all the houses you can see. It was really exciting. Honestly, it was a modern housing estate in 1790. They have the date of the brick up there. It was built in 1780. The point is that William Morton Pitt went around Corfe Castle and collected individual-level data on income. The world was pretty straightforward then. People did not move very much; people did not change very much. People answered the questions, particularly if the lord of the manor turned up.

We are in a completely different world now. People need data quickly. The local authorities that we talk to recognise that, if you just do a census and then do another census 10 years later, the in-between estimates get progressively worse over time. You are having to adjust for internal migration and international migration.

We do not have a system in our country that says, "If you move house, you have to tell someone". For example, we use GP data as a proxy for some kinds of internal migration, but, of course, particularly young men do not register with somewhere new when they move.

It really is a challenge, and we recognise that challenge. What we have been doing is saying, "Are there a variety of administrative data sources that we could bring together on a regular basis and produce local area population estimates?" We have shown that we can do that. We can do it



HOUSE OF COMMONS

with very close accuracy to the census. Remember, the 2021 census was probably the best one we have ever done in England and Wales.

We can do it with very close accuracy. We can do it much faster than we used to do them previously, and we can do them in a way that enables local authorities to bring those data in very quickly. So far, so good.

We have also demonstrated that we can get income using HMRC data to a very local level. It absolutely needs to be privacy-enhanced; do not get me wrong. We can give you the average income of very small areas. That is something that users want. We have also demonstrated that we can do multivariate work, such as ethnicity by income, at a local level. Again, that is privacy-enhanced.

We have demonstrated that we can answer all the questions we were asked by the Government in 2014 when they said, "Okay, let us do a census in 2021, but can you work on a programme to see if you can do administrative data, can those comparisons be made, and can you do income?" I believe we have answered all the exam questions that we were set in 2014, using the 2021 census data as comparison. That is why we have now gone out on a consultation; we launched a consultation at the end of June, as you know, and it is running through until the end of October.

We have had around 200 responses so far, many of them individuals, and a number of them, of course, from genealogists. I am sure we have all looked at the 1891 or the 1911 census to see our forebears. Let me also say that we have a piece in the consultation that explains exactly how, in 100 years' time, people will be able to find out broadly what we were like—not what we were doing, but where we were living and what our household arrangements were and things like that.

We still continue to work on one or two areas, particularly occupation, and this comes back to my point about administrative data. You may think, "Surely somebody collects occupation". It is collected in some ways, and we are making great progress, but HMRC does not collect occupation. It did a consultation earlier this year as to whether occupation should be added to the self-assessments, but the view was not to do that, which I was disappointed with, I must be honest, but that is the decision that they have made. Would you like me to run through the timetables?

Q28 Karin Smyth: I was about to ask you. When is the consultation window and when do you expect to inform Ministers?

Professor Sir Ian Diamond: The plan is that the consultation ends at the end of October. My recommendation is to the board of UKSA. I intend on making that recommendation at the December board meeting. It will be then for Robert Chote as the chair to decide how to publish that, but I suspect the recommendation from me to the board is not going to stay private for very long. There will then be a short period of time, probably



HOUSE OF COMMONS

six to eight weeks or something like that, when we will actually have to get the formal writing, translation into Welsh and things like that done. I imagine that will be the documentation that goes around a recommendation that will be known, I suspect.

Q29 **Lloyd Russell-Moyle:** While the ONS is only responsible for England and Wales, you as the National Statistician have that broader responsibility for the whole of the UK. How are you considering the whole of the UK approach in these discussions around the future of the census?

Professor Sir Ian Diamond: That is a really important point. I will come to Wales and Northern Ireland in just a moment, because I may start with where we have not made an enormous amount of progress so far. That is in Scotland. The reason for that, as colleagues will know, is that Scotland delayed its census, as it is entirely allowed to do, and has had to work very hard on its under-enumeration estimation. It will announce its first results on 14 September.

Therefore, I do not think—rightly, for the amount of work they have done, which is a huge amount of work—in Scotland there has been an enormous consideration about the next stage, and now there will be. We will support that in exactly the same way as we have provided enormous amount of support to our colleagues in Scotland to enable them to produce the first results from the census.

Q30 **Lloyd Russell-Moyle:** Is your view that the UK should move together on any near future census, or is it acceptable for there to be different approaches in different nations?

Professor Sir Ian Diamond: Ideally, we should all move together. That is my view, but I recognise that that is a view. With regard to Wales and Northern Ireland, we have an enormous amount of work with them. Two of my senior colleagues on population were in Northern Ireland last week talking about the issues, and we have made it very clear that we will provide any support necessary in what we recognise is an independent decision that they will need to make, and similarly with our colleagues in Wales. Again, we have spent a lot of time with them and we will continue to spend time with them, but at the end of the day it is their decision and they will make it.

Q31 **Lloyd Russell-Moyle:** Are there any bits of data, moving away from the census, that you would struggle to find using administrative data or even enhanced administrative data? Clearly for professions you might be able to persuade HMRC to reconsider or make a tweak, but I am thinking particularly of some of those personal questions like religion, sexuality, et cetera. There seems to be less centralised reporting, effectively, of those.

Professor Sir Ian Diamond: You are entirely right. It should be possible to make some progress around religion, but we are honest. When it comes to sexual identity and sexual preference, that is an area in which we are working with our users to understand their needs and then to work out how best to meet those needs.



Q32 **Lloyd Russell-Moyle:** Is it the case that some of those things just will not be recorded in future because there is not actually a use for it beyond interest—a bit of voyeurism but actually it is not really used for service delivery—or is it that there is a great need for that and you will need to find alternative sources?

Professor Sir Ian Diamond: That is for users to tell me, to be honest. I can assure you that we are and will be working with a wide range of users to understand the user need for particular questions and then we will meet those user needs as best we can, but also to recognise that the geography of the user need may vary. Do we need those data at a very local level, or are users who are needing them for services and policy purposes happier to have them at a higher level of aggregation, in which case the estimation can vary?

Q33 **Lloyd Russell-Moyle:** You expect that these proposals would potentially cut the cost in about half of a potential 2031 census. How did you get to those estimates?

Professor Sir Ian Diamond: We are in the process of really working out what the economic case is. There is a lot of work going on. The initial estimates were basically to say, "We have calculated pretty much what the economic benefit is of the population estimates in the year after the census". The benefit of those estimates goes down over a period of time as they become less accurate, particularly, we acknowledge, at a local area level. Let us assume that we can make those accurate estimates each time. You can imagine that you are then able to estimate the area between those two lines, and that gives you the economic benefit at one level, which is the economic benefit of accurate data on an annual basis instead of having to wait for 10 years.

The second thing is just to estimate the overall cost of doing the census. The one thing that is an absolute given is that inflation impacts on censuses. I suspect the 1801 census did not cost an enormous amount of money. A 2031 census is in the billions. Then, being able to work out what it would cost to do administrative-based estimates, we can make that distinction.

We are now working very carefully on what we think the costs would be over the future of doing the administrative-based estimates. Some of that is working with other Government Departments to make sure that we can get the regular flows of data that we need. To come back to the question that I answered right at the beginning, those data need engineering, to use the technical term, to make sure that they are cleaned and to make sure that they are in the shape that we need them to be able to use.

When I wrote to each permanent secretary to advise them of the consultation I said, "And these are the data that we will need from you on a regular basis". Working out the costs of getting those data is something we are working on with Government Departments now as we make the final business case.



Q34 **Lloyd Russell-Moyle:** It may well be that there is not a huge overall saving in the end in terms of costs if you are having to then invest in alternative sources of data and other Government Departments having to pay for their data to be passed on to you, but it has an economic benefit that is much wider. Is that possibly what you are saying?

Professor Sir Ian Diamond: My own personal view when we get there is that the actual cost will be rather less. This is a win-win, where I am coming from, although let me be clear that we have not made a decision yet; this is a real consultation. The potential win-win is that we will be able to collect the data and to produce the data cheaper, but at the same time the economic benefit will be much greater because local authorities, for example, and other service users will have more accurate data more frequently.

Q35 **Lloyd Russell-Moyle:** Do you think you will publish that data and analysis around costs that you are drawing down when you make your recommendations?

Professor Sir Ian Diamond: Yes.

Q36 **Lloyd Russell-Moyle:** It will be that timescale. Is there an option here for a hybrid version? You move to more administrative data but you still do some blanket surveys maybe not on even a 10-year basis but on a periodic basis?

Professor Sir Ian Diamond: There is a real case for ensuring over time that you have some benchmarking. It is really important. Now, whether those benchmarks can come from some administrative data or whether you need a large-scale short data collection exercise is something we are looking at at the moment.

Q37 **Lloyd Russell-Moyle:** Some of the censuses in the early 1900s combined electoral rolls and census. That has been split. Is there a case to bring some of that together so that that allows your benchmarking?

Professor Sir Ian Diamond: There is a certain opportunity there. Of course, if we were in Scandinavia this would not be an issue because we would have a population register. I recognise that is not something that in the United Kingdom we have been prepared to engage with; that is fine, but there is a situation where you use just administrative data to get real-time information. I say that because I think there are different methods that you could use, and that benchmarking is going to be important.

Q38 **Lloyd Russell-Moyle:** Forming the electoral roll into a population roll could be a neat solution but whether there is political will for it is a different question.

Professor Sir Ian Diamond: The ethics would be quite interesting. Therefore, it is not one that I have proposed nor will it be in the recommendations.



Q39 **Damien Moore:** There appears to be a universal agreement that data are not being shared effectively between Government Departments in support of statistics and research. First of all, why is this? With all this talk of administrative data, I am tempted to ask the question of whether it would be better if there was a Department looking after this, perhaps a Department for Administrative Affairs?

Professor Sir Ian Diamond: A Department for Administrative Affairs could be doing an awful lot of things. I concur that data are not being shared as effectively as they should. No one could say that. The first point to make is there is no legal impediment that prevents it. We have all the legal issues. I might like to have health data in the Digital Economy Act, and certainly that is something that Professor Cathie Sudlow, who is doing a review of the availability of health data at the moment, will be considering, I think, although I do not know.

There is no legal impediment to sharing data, so why are data not shared? There is a real issue around the fact that we are a Government of Departments that operate almost, in many ways, as independent entities, and that therefore people are rightly, given all these cybersecurity issues, nervous about sharing data. There is an inherent nervousness and an inherent concern that no one wishes to be the people who enable public data to be at risk.

Therefore, if we are going to make progress, I come not necessarily to the Department for Administrative Affairs, but a number of very sensible suggestions have been made. Cassie Smith, who is at Health Data Research UK, has proposed that the Treasury might say to all Departments, "As long as it is done properly and against the right security constraints, we will back you".

There is also the case that you could move to one data owner. There are two problems at the moment. One is that you have separate data owners everywhere, and so you have to negotiate bilateral agreements to do anything. Secondly, data owners say, "You can use those data for purpose X", and then, if you want to use them for a slightly different purpose, you have to go back through the process again.

Enabling data sharing to happen is one of the most important things that we could do. We did it during the pandemic. We understood the dreadful inequalities by ethnicity that came around covid mortality. We were able to do that by linking data from different sources and controlling for disadvantage and things like that. It was really sad work but incredibly useful. We were able to do that because during the pandemic there was a special dispensation that allowed data to be shared. They were called COPI notices.

All the work that we were able to do around, sadly, covid mortality or around, for example, long covid, we could not do for cardiovascular disease; it was only for pandemic insights. Of course, that has stopped now; you simply cannot do all the things you might want to do around



HOUSE OF COMMONS

what is going on in cardiovascular disease or cancer. Therefore, one data owner would be a very good way.

It is also worth saying that, when people talk about data sharing, they believe that there is going to be some enormous lake of data and that people are going to be marching around with different datasets. That is absolutely not the case. Data can stay exactly where they are held, in an entirely safe environment, and we are able then to access the bits that we want in a secure way.

At the ONS we have been privileged to lead a cross-Government piece of work to lead up to a multi-cloud platform called the Integrated Data Service. That now works, but it requires that we get many more datasets. The speed of upskilling what will be a transformational programme for Government will come through really improving the data ownership and the sharing of data.

Q40 Damien Moore: Thank you. That is really useful. Just touching on some of the points you raised earlier, how well does the existing legislative framework provide for access to data for statisticians in the devolved Administrations, because we have added that element of complexity?

Professor Sir Ian Diamond: It is an element of complexity. Certainly we try to help. There are trusted research environments in each of the three Administrations, and we try to work with those trusted research environments, but it is the case. One example is that we are able to have data from HMRC, which I have talked about earlier, on payroll numbers and things. Our colleagues in the Northern Ireland Statistics and Research Agency have to negotiate their own arrangements. I cannot say, "I have them so I will let you have them". It is very difficult in that way.

We need to recognise that our colleagues in NISRA, as it is called, have challenges of resource. They do not have lots of people. I do not have lots of people. None of us has lots of people running around negotiating data agreements, but I have a few more people than they have. It is a real challenge.

Damien Moore: That is something we definitely need to focus on, then.

Professor Sir Ian Diamond: I think so.

Q41 John Stevenson: You have just touched upon the challenges of sharing data between Departments. What about the challenges of statisticians and analysts getting access to Government data in a timely fashion? What are the challenges there?

Professor Sir Ian Diamond: It comes down to the whole issue of getting those data to be enabled into a trusted research environment where they can be analysed and where data are being shared. There are two questions there. First, do we need to link data? If we need to link data, in the main, we can do that quite quickly, except—and here comes



the big “except”—because there are different infrastructures in different Departments, data do not always equally easily talk to each other. That is why I mentioned some of the issues around data engineering. Different Departments will have different systems. There is an enormous amount of work that can make it very difficult to share data.

Q42 John Stevenson: That is sharing data. What about statisticians actually getting access to that data?

Professor Sir Ian Diamond: I am terribly sorry; when I was talking about sharing data, I was meaning either an individual dataset or data that had to be linked. With regard to statisticians having access to those data, again, that comes down to the data owner being prepared to give access. That varies. To return to something I said earlier, we do it according to a set of principles: that the work is ethical, that the work is in the public interest, that the work can be done—the statistician has the skills to do it—that what is going on is transparent and that the data are held very securely.

We do it according to those sets of principles, but there is something that I would want just to mention. Sometimes data owners give you the permission to do one particular piece of analysis, but statistical analysis does not actually quite work like that. You do not go in and randomly, just say, “Oh, I wonder what is here”. You are asking particular questions, but you often want to go and explore the data and look in a very imaginative way. That can sometimes be difficult if your permissions only allow you to do a very tightly defined model. Enabling exploratory data analysis seems to me to be incredibly important.

Q43 John Stevenson: Is the Central Digital and Data Office’s data sharing initiative helping in any way?

Professor Sir Ian Diamond: Yes, I am a huge fan. This will bring the data and make them available, and the more datasets that are there, the more other people will say, “I really ought to be there because everybody else is there”. There is a real opportunity for that data marketplace, as it is called, to really be a deal-breaker.

Q44 John Stevenson: Given the issues we have just discussed, are you confident you are going to make robust recommendations to Government regarding a shift away from the present census routine into more admin data in the near future?

Professor Sir Ian Diamond: I am conscious I am going to make a robust recommendation. I hope I have the privilege of coming here for a conversation, but I do not want any of you to say, “I do not think this is terribly robust”, so I am conscious that whatever recommendation I make will be robust. We are doing a consultation. Let me assure you that William will get a text as soon as we have a recommendation that he can share with you.

Q45 John Stevenson: Do you think the Government Departments will be



receptive?

Professor Sir Ian Diamond: They will. The reason that they will is because my colleagues have put so much effort into engagement at all levels. This consultation has not been one where we have just put something on a website and we will wait to see what comes in. They have been out and about talking to people. When the recommendation comes, I do not think it will be a great surprise.

As I indicated, we have been asking people, "What do you want? What are the key things you would want to see?" If we cannot deliver what people want then you know what the recommendation is going to be. In addition, as I said, I have written to all Departments and said, "These are the datasets that we are going to need from you on a regular basis. Let us just be honest about that now, and if there is a problem, we need to have a conversation now rather than later".

Chair: I am looking for that robust text in due course.

Professor Sir Ian Diamond: I will put "robust" on the top.

Q46 **Chair:** We appreciate that. There are a couple of quick questions from me to finish our session today, Sir Ian. You have outlined in quite some detail the experience in the UK of integrating new data sources and so on. Do you have any quick comparisons to make to other countries around the world of where we are at and where they might be?

Professor Sir Ian Diamond: We are ahead of an awful lot of countries, but other countries are also making real progress. The Dutch have been making a lot of interesting progress. My counterpart in Canada is coming over in October to talk to me about some of the things that they are doing, because he wants to learn from us and we want to learn from them. Colleagues in the Antipodes have made some good progress.

Q47 **Chair:** Just in terms of the census, have any other nations transitioned away from that model in the way your proposals recommend that the ONS does, and what has their experience been?

Professor Sir Ian Diamond: I would start by saying I have not yet made a recommendation. A number of countries have moved away from a census but into different models. The French moved to a system whereby they have a series of large biannual surveys that they link together. The journey that we are on is similar to that which other countries are on, driven in some cases not only by the availability of administrative data and not only by the need for more frequent data, but also by, in some countries, the challenges of doing a census where overall enumeration rates have been somewhat lower than ideally we would like.

Nobody is 10 years ahead of us, and therefore I can report what they are doing, but there are a number of other countries that we are talking very



HOUSE OF COMMONS

closely to and that I suspect will not do censuses in a similar way in the 2030s.

Chair: Sir Ian, we are very grateful for your attendance at the Committee, as ever. Thank you for your insights. If there is anything you wish to illuminate further, please do write or indeed send urgent texts. I would just thank you on behalf of the Committee this morning.

Professor Sir Ian Diamond: I hope that it has been as enjoyable for you as it has for me. It is an unbelievable privilege to be here. I thought the questions were challenging and incisive, and I hope that my responses were reasoned.

Chair: You can certainly come again if that is your review. We are very grateful. Sir Ian, thank you very much.