

# Public Accounts Committee

## Oral evidence: Use of evaluation and financial modelling, HC 1055

Wednesday 9 March 2022

Ordered by the House of Commons to be published on 9 March 2022.

[Watch the meeting](#)

Members present: Dame Meg Hillier (Chair); Sir Geoffrey Clifton-Brown, Peter Grant; Kate Green; Sarah Olney.

Gareth Davies, Comptroller and Auditor General, and David Fairbrother, Treasury Officer of Accounts, were in attendance.

Questions 1-73

### Witnesses

**I:** Sir Tom Scholar, Permanent Secretary, Her Majesty's Treasury, Catherine Hutchinson, Head of the Evaluation Task Force, Cabinet Office, Professor Sir Ian Diamond, UK National Statistician and Head of the Government Statistical Service and the Government Analysis Function, UK Statistics Authority, and Cat Little, Head of the Government Finance Function, Her Majesty's Treasury.



Report by the Comptroller and Auditor General  
Evaluating government spending (HC 860)

Examination of witnesses

Witnesses: Sir Tom Scholar, Catherine Hutchinson, Professor Sir Ian Diamond and Cat Little.

**Chair:** Welcome to the Public Accounts Committee on Wednesday 9 March 2022. Today we will be looking at how Government models situations and evaluates programmes. Modelling has been very much in the public eye as a result of the pandemic, but Government have used modelling for many years. We want to look in depth—thanks to the National Audit Office and its Reports—at how modelling is worked and used through Government, and at how programmes are evaluated to see how well they have worked. That is vital for politicians, but it is perhaps often underplayed.

We have before us some top officials from the Treasury and wider Government: Sir Tom Scholar, the permanent secretary at the Treasury, and Catherine Hutchinson, head of the evaluation taskforce at the Cabinet Office. I am pleased also to welcome Professor Sir Ian Diamond, the UK national statistician, and head of the Government Statistics Service and the Government Analysis Function at the UK Statistics Authority. That is a very long title that I think underlines the important work you do, Professor Diamond—we are keen to hear from you. Welcome to those of you who are new to the Committee. Welcome back to Cat Little, head the Government Finance Function at the Treasury, among other roles.

This subject is a matter of great interest for the Committee. We obviously look at data a lot—we need good data to have good evaluation—and we are keen to see how that is going and see the focus at the centre of Government on this issue. I will ask Peter Grant MP to kick off.

Q1 **Peter Grant:** Thank you, Chair, and good afternoon to our witnesses. Sir Tom, the NAO Report on the evaluation of Government spending found that in 2019, of £432 billion of planned expenditure, only £35 billion had plans for sufficiently robust evaluation. How can you or anybody know that Government spending is making a difference if nobody evaluates it afterwards?

**Sir Tom Scholar:** The short answer to the question—there is a longer answer, which I am sure we will get into—is that each Department is responsible for ensuring the effectiveness of its own spending. Each Department has an accounting officer who is directly accountable to Parliament for that, and you obviously see accounting officers all the time at this Committee.



Evaluation, as we are certainly discussing this afternoon, is an extremely important part of ensuring the effectiveness of public spending, the quality of public services, value for money and so on. It is not the only tool that Departments have; it is a very important tool, and our job is to promote it and improve its use. But ultimately it is the accounting officer who is responsible and it is the role of Parliament to hold accounting officers to those responsibilities, and to account.

**Q2 Peter Grant:** But as permanent secretary to the Treasury, you have got a significant role in deciding which of these projects are going to be approved for taxpayer funding. Why do you not insist on adequate evaluation plans, either before you approve them or before you give recommendations to Ministers to approve them?

**Sir Tom Scholar:** As the Report says, the last couple of years have seen a real increase in the emphasis that we put on evaluation. I think I should stress that the particular role of the Treasury in evaluation comes when either assessing new proposals for spending or setting spending plans for a Department. That typically happens at a spending review.

We had a spending review in 2020 and 2021. Prior to that, we hadn't had one since 2015, so I think one reason why there has been a big step up in the activity has been because the Treasury, the whole of the centre of Government, has been making big, strategic and long-term decisions in those two spending reviews.

We did very much increase the emphasis on evaluation in the 2021 spending review. I am sure Cat will talk more about this, but for the first time we made commitments on improving evaluation, on publishing evaluation strategies. We made those commitments part of the settlement with spending, so that is a condition of the funding being provided. We have got a whole set of mechanisms for following up on that, and ultimately holding Departments to account for it.

What I would say is that we recognise the picture in the Report: that the Government need to do more to strengthen the bite, if I may put it like that, and the traction of evaluation; and we are trying to do that through things such as setting up the evaluation taskforce just less a year ago and the new actions that we have taken on the spending review.

**Q3 Peter Grant:** Have you been given a clear steer from your Minister, the Chancellor, that he expects this to be done? Is it something that has been driven by Ministers as well as by yourselves?

**Sir Tom Scholar:** Absolutely, yes—indeed. The focus on evaluation was all part of our spending review preparations and that was fully done with Ministers and at their direction; and in the run-up to the spending review last year the Chief Secretary to the Treasury and the Chancellor of the Duchy of Lancaster wrote to all heads of Department, setting out that this was a big priority for them. If you look more generally at statements by senior Ministers, Ministers and publications by the Government about Government reform, you will see a consistent emphasis there on



## HOUSE OF COMMONS

improving the quality of data. Evidence-based policy making and evaluation was part of that.

Q4 **Peter Grant:** Why has it taken so long? What is the difficulty in introducing this as standard defined practice across Government?

**Sir Tom Scholar:** I think it is fair to say that evaluation has always been an important tool, which—

Q5 **Peter Grant:** If it is an important tool, why is it that three years ago, only 8% of major projects had any plans for a proper evaluation? It does not appear to me that your colleagues in other Departments see it as particularly important.

**Sir Tom Scholar:** As I said earlier, the first standards were published seven or eight years ago. These things are extremely complicated to do in some cases, and you need to work hard over time at improving performance. As I said earlier, I think the two spending reviews in the last two years have had a real galvanising effect. We are now seeing greater traction in Departments, but I certainly acknowledge that we have a lot more to do.

Q6 **Peter Grant:** Thank you. Ms Hutchison, can you give me a couple of examples of practical things that you are doing in order to get this established in the way that it is quite obvious we all want?

**Catherine Hutchinson:** Absolutely. The evaluation taskforce was set up in response to some of the figures that were discovered. Put simply, just the existence of a taskforce is a strong practical example of something that the Government has already started to do.

Some of the things that the evaluation taskforce has delivered have been that we have ensured, working with Treasury partners, that we have three key evaluation requirements in every Department's settlement conditions, which enables us to much more strongly follow up with Departments about the existence of strategies—what have they published and are saying they are going to do for evaluations across the whole Department?—and about their evidence gaps, because we understand that there are significant gaps. The NAO Report has highlighted those, but to understand what those gaps are, we have to start from somewhere.

Finally, as Sir Tom pointed out, the spending review is mostly about new spend, so what about all the existing programmes that are put in place? For that, we have asked them for the strength of the evidence that exists for their 20 largest spends. They will be required to tell the evaluation taskforce those three things. In addition, we have specific requirements in all the settlement conditions for the programmes and projects that we think are the most important.

Q7 **Peter Grant:** Thank you. I will go back to Sir Tom for my final question for now—just to set out why, for me as a Member of Parliament, this is important. We are looking at £432 billion of expenditure. At the end of it, and possibly all the way through it, politicians on one side of the House will say, "This is wonderful. We are wonderful. Look at all the good we



## HOUSE OF COMMONS

are doing.” Politicians on the other side of the House will say, “You’re spending lots of money. You’re wasting money, and it is not making any difference to people’s lives.” How are the poor souls out there supposed to judge those two competing views when it comes down for them to decide whether to re-elect the Government or to seek a change?

**Sir Tom Scholar:** The first thing I would say on that is that the formal evaluation of a programme within a Government Department is an extremely important part of designing a programme, implementing it and then making decisions on whether to extend, change or possibly end it. But it is part of a much broader context of evaluating and assessing the success or otherwise of Government intervention.

This is a highly crowded market, with think-tanks, universities, Select Committees, parliamentarians and journalists. Many people are commenting, often from positions of great expertise and authority, on the interventions of Government, so the full evaluation within a Department—important though it is—will only ever be a small part of that landscape. That said, the strong view of the Departments represented here this afternoon is that we need both more formal evaluation from Government Departments and greater transparency around it, precisely for the reason that you set out.

As Catherine has said, we are now, for the first time, requiring Departments to publish an evaluation strategy. Any decent evaluation strategy will have a communication strategy as part of it, because communication is a central part of evaluation. I really hope that through this year and over the years to come, formal Government evaluation will be more visible and, therefore, a bigger part of that overall policy debate.

Q8 **Sarah Olney:** In the financial modelling Report, quite a lot of review has been done of various models being used across Government. Some of them play quite a fundamental role in both reporting and accounts, and yet errors have been found. I am thinking particularly of paragraph 2.6, for example, which mentions errors of £800 million and £45 million in models that were used to produce estimates for financial accounts.

Are you concerned that sometimes these models are being used when they have not been properly reviewed by an independent party—someone who was not involved in putting together the original model?

**Cat Little:** *indicated assent.*

**Sarah Olney:** I see you nodding.

**Cat Little:** Yes, is the simple answer. In the finance function, I think the areas you refer to relate to models that underpin the financial reporting of a Department. There is a very simple “three lines of defence” assurance model that we expect for all significant models that we use both for financial management and for financial reporting.

In the first line, we set out very specifically what we expect finance and analyst users to do in order to check the sources, to check the most sensitive assumptions that are within the model, and to check source data



to make sure that we have checked all the links and that the macros work. In the second line, we expect there to be some sort of independent specialist who actually checks models. In the third line, we should be using our internal auditors as well as our external auditors to make sure that we have got a really good assurance framework. That is obviously the backstop for what you would expect.

It is always disappointing when you find errors in the third line of defence, because it means that something has probably gone wrong in the first and second lines.

**Q9 Sarah Olney:** Is that your understanding of what has happened in the two particular examples that have been quoted?

**Cat Little:** I have spoken to the finance director, and I am conscious that the Department is not named and there are very specific things in there. My understanding is that with the error that was found, while it was part of the audit process, the finance director had done some initial quality assurance that led to this being identified as part of the audit process. Of course, that is too late.

The other thing I would say about financial reporting is that we have a certain number of accounting standards that require us to set out how we assure the most sensitive judgments within our financial reports. It is a core part of the accounting standards that we use. We should be checking sensitivities and setting out precisely which models have the most material impact on the financial statement. In this case, I think that was well understood; it was just a matter of the model having an error in it.

**Q10 Sarah Olney:** Sir Ian, do you think that is sufficient in terms of independent review of models being used—particularly, as I say, where it is material to accounts and reporting?

**Professor Sir Ian Diamond:** I am passionate about the “three lines of defence” approach, which does mean that we have a responsibility, both initially to define the model correctly and—I stress this—equally importantly to make sure that the data that are going into those models are absolutely accurate.

You will, I am sure, have heard of the famous phrase “Garbage in, garbage out”, and that is absolutely the case. Part of that first line has to be, “Are we using the best data, are there any potential errors in those data and what is the uncertainty in those data?” Then we can estimate the model. Again, models come in different ways.

I think we have a pretty good system where those three lines of defence are absolutely there, but I do think, and this is one of the things that the analysis function is clear on, that if the departmental director of analysis, or indeed the finance officer, is not sure—maybe the person is using a very complex model, so let’s get expertise elsewhere—we always stand ready to help.



## HOUSE OF COMMONS

On your question about whether that is enough independent review, I think the third line of defence has to be, and is, independent. I think that is important, but we need to build a culture of accuracy inside the Department first of all.

- Q11 **Sarah Olney:** Thank you. Sir Tom, just to pick up on that point about uncertainty, do you think Departments are doing enough to include and report uncertainty properly in their models?

**Sir Tom Scholar:** In an uncertain world, any model that does not explicitly address the uncertainty is deficient.

As I think the Report says, we don't check every model that every Department develops, because they develop many for their own purposes, but we would certainly routinely interrogate any model used to generate analysis to support a case for additional spending or any other case that comes to the Treasury in a policy sense.

The first question is, "What are your assumptions about this? What are your assumptions about that? What is your range? That is your central case, but what is your confidence interval?" And so on and so forth. Dealing with uncertainty properly is fundamental to decent modelling, and that would be driven partly by the standards that we set, partly by the skills and capability building that Ian is doing through the analysis function and then certainly through the cross-examination of models and their results that we do in the centre, whether that is done by the Treasury or the evaluation task force. It is obviously central to quality modelling.

**Professor Sir Ian Diamond:** Could I add that it is absolutely essential? Just this morning, I gave a data masterclass to a number of Ministers, and one of the things I said was, "You must demand, for any data you see, that the uncertainty is demonstrated and pictured." There are now very good examples of how to do this. The Bank of England do it extremely well, because they don't even give, for the future, a central estimate; they simply provide a range of uncertainty. I think that is incredibly helpful. I think it is important to distinguish between different types of uncertainty. Some uncertainty you can actually measure quite accurately; with others, you are simulating. But you need to be absolutely clear, as Tom says, about the assumptions and you need to display the uncertainty. That's critical.

- Q12 **Chair:** Data masterclasses for Ministers sound great. It may be a bit Yes *Minister* for some people, but is that something you are doing regularly?

**Professor Sir Ian Diamond:** Absolutely. I did one on Monday this week for the Public Administration and Constitutional Affairs Committee. If this Committee would like to spend an hour doing that—

**Chair:** We are the sort of Committee that would love to do this.

**Professor Sir Ian Diamond:** We would be delighted.

- Q13 **Chair:** How many Ministers are you getting round?



## HOUSE OF COMMONS

**Professor Sir Ian Diamond:** This morning we had eight. I would say around half the Public Administration and Constitutional Affairs Committee were visiting Wales, where the ONS office was, and they all came and it was very good. We will continue to do this. It is very enjoyable. It is interactive and—

Q14 **Chair:** We don't need you to give the advert; we're sold. The question is really how many Ministers are doing it. That is pretty critical, because they are ultimately going to be driving this.

**Professor Sir Ian Diamond:** There were Ministers of State and Secretaries of State there this morning.

**Chair:** Secretaries of State—that is very impressive; well, it's a start, I suppose, but that is not your problem. Thank you for that; it is helpful. Ms Olney.

Q15 **Sarah Olney:** What more is being done to embed that understanding of the importance of uncertainty in modelling beyond Ministers and Secretaries of State to officials and to people working on the models in the offices?

**Professor Sir Ian Diamond:** I will be clear that it is a work in progress, but most permanent secretaries and many senior civil servants have undertaken an online, five-module data masterclass. I did one of the modules, and others have done that. That has been very well received. We are now extending that into an evaluation masterclass.

We are also talking right across the analysis function as we understand what the standards are that are needed to make sure everyone in the analysis function has those standards. Critically, Tamara Finkelstein, the permanent secretary of DEFRA, leads the policy profession and she has been brilliant in saying that we need to understand and build quantitative skills in the policy profession.

Lord Maude, in his report recently, suggested that we should take an audit of quantitative skills in the policy profession. We have now done that; we haven't quite finished the analysis. Critically, what that means is that in future, I hope, we will have, from day zero, analysts and policy professionals working together, where the conversation can be one of equals in terms of the quantitative skills. That is super-important.

Q16 **Sarah Olney:** I don't know if this is a question for you or one of the other witnesses, but on the role of the analysis function, is that to ensure that we have some common standards in modelling across different Departments?

**Professor Sir Ian Diamond:** Certainly. The analysis function, which was only set up a few years ago, includes all Government economists, statisticians, operational researchers, social researchers, actuaries, geographers and most data scientists. The idea was that many of the skills that people are using actually overlap.



## HOUSE OF COMMONS

While recognising the particular professions, we are now in the process of building a set of standards that include modelling, which will be held right across Government. We have a workshop in March: here we are, this is where we are, and we will now move them forward. There will be an expectation that departmental directors of analysis own these standards and report back on an annual basis to those standards.

Alongside that, we will have an identification of where there are particular centres of excellence. If you are, for example, thinking about a system, working towards net zero might be a system that needs a multidisciplinary approach. That system I would call a hierarchical chain graph. You need to estimate that hierarchical chain graph. I would love to think that we needed 30,000 people across the civil service who could estimate hierarchical chain graphs. We don't, so we need to identify where those centres of excellence are and enable all Departments to be able to draw on them so that we raise overall the standard of modelling.

**Cat Little:** Could I answer in respect of the finance function? As Ian and I were discussing yesterday, we have mutually aligned responsibilities for standards. The finance function for financial modelling has a series of activities that we are undertaking.

We have a finance standard and have done for the last three years. We also have a self-assessment tool that allows Departments to assess whether they are meeting the standard, which we assess on an annual basis. In particular, we have been focusing on forecasting because that is where we have had the most difficulties and the most inconsistency in the standards that Departments are using.

Last year we implemented standard MI reporting for all boards, which has the financial risks and opportunities and therefore how you estimate financial uncertainty within your forecasting at a departmental level. We are also going to update the standards to set out in more detail the sorts of skills, behaviours and the MI that Departments should use to try and raise standards overall.

Q17 **Sarah Olney:** Excellent. Sir Tom, do you expect to see these standards around uncertainty reflected in the reports to you in future? Is that what you will be demanding from Departments?

**Sir Tom Scholar:** We already expect that. I am talking particularly here about bids or proposals for new areas of Government spending. We already expect uncertainty to be fully incorporated into the modelling. As I said earlier, more or less our first question is to check that out. More generally, through that exercise, we build an assessment of the degree to which any particular Department has understood that it has got a good quality function and is able to do that. That then helps us reach a risk assessment as to how we approach these things, including through spending reviews.

What we really hope—this is I think a collective hope—is that a sustained drive to improve standards, not least now that we have got the analytical

function properly established, will, as Ian was saying, over time improve standards across the board. After all, as the Report says, it was the consequence of a very serious modelling error back in 2012 in respect of awarding the franchise for the west coast main line. In effect, it was a fairly basic modelling error, which then had very serious real-world consequences. It was in dealing with that that my predecessor, Nick Macpherson, wrote his report on modelling in Government. In a sense, that is where this whole effort began. As Ian said, it has some way to go, but we definitely expect to see improved standards.

- Q18 **Sarah Olney:** Thank you. In terms of upholding those modelling and evaluation standards and driving improvements, who is ultimately responsible for making that happen? Who is the person in charge of delivering that?

**Professor Sir Ian Diamond:** There is an overlap between financial modelling and statistical modelling, but there are also specific elements. Between Cat and me, we would expect to ultimately own that, but as I indicated earlier, the ownership is within Departments, and the analysis function has a very strong network of departmental directors of analysis.

The future expectation is that the departmental director of analysis would own the standards and would report on an annual basis about how those standards have been upheld. At the same time, we will maintain really strong continual professional improvement. Some of the things that you would be able to do on your mobile phone while talking to me now would have taken me three weeks 30 years ago. We need to recognise that we have to be at the cutting edge of modelling in Government, and not lagging behind. That is another thing that the function will be taking forward, making sure that we are upskilling people on a regular basis.

**Cat Little:** Going back to the role of the accounting officer, every accounting officer has to make sure that they have an assurance framework in place that gives them adequate assurance on the financial models that are business critical to their operation. In addition to the functional role that we play in making sure that people raise their standards, we need to make sure that accounting officers are complying with their requirements. "Managing Public Money" sets out two things: first, that accounting officers should comply with the functional standards and, secondly, that financial modelling is an important part of the responsibility of the accounting officer.

- Q19 **Sarah Olney:** You have a departmental director of analysis in each Department, but who is holding them accountable for their performance, and how are you benchmarking performance across Departments?

**Professor Sir Ian Diamond:** There is the classic horizontal and vertical model. Vertically, the accounting officer will have a line management responsibility through to the departmental director of analysis, but horizontally it comes across to me. Certainly, we, as the analysis function, would expect to be involved in conversations at any time about this. If



## HOUSE OF COMMONS

someone felt that they were under undue pressure or whatever, again the analysis function has a role to play.

Q20 **Sarah Olney:** Going back to evaluation requirements as part of the spending review settlement conditions, how do you know that Departments are complying with those evaluation requirements if you are not following up?

**Sir Tom Scholar:** We are following up, and Cat and possibly Catherine can explain how.

Q21 **Chair:** The two Catherines! People are trembling out in Whitehall.

**Cat Little:** We have a dominance of Catherines—it is great. The three conditions that Catherine alluded to earlier have very specific dates in the settlement letters, which require follow-up action between the Treasury and the evaluation taskforce. Those three items should all be concluded by July this year, and we are currently following up with Departments and making sure that they are on track. More broadly, all conditions, both within settlement letters for fiscal events and for any additional spending that we sign off or give Treasury consent for, are tracked by each spending team.

One of the things we have been working on this year is that there have been inconsistent tracking tools across the Treasury, depending on the spending team, so we have been looking at how we can standardise that and keep a running log of all the conditions, whether they have been met and whether they were done on time. Certainly for the Departments with the most conditions, like the Department of Health, we have regular conversations about why they have not been met. As this Committee knows, there are some examples where conditions were not met, which have led to follow-up action.

Q22 **Sarah Olney:** Can I talk about some of the examples? Obviously, we are exposed to those through the course of our work. On courts and tribunals, for example, 127 courts have been closed since 2015, but there is no formal evaluation of the impact of that. However, as constituency MPs, we all see the impact on our constituents of being denied access to justice. Why has there not been formal evaluation of the impact? Is that as a result of processes that have been changed?

**Cat Little:** If I could, I will talk more widely about how we go about enforcing evaluation of major Government projects and programmes. The closures started ahead of the court reform programme, but for all intents and purposes we now expect it to be wrapped up within that major Government programme.

The IPA, as part of their stage-gate review processes, have a standard methodology, which includes the assessment of whether there are evaluation strategies and whether they have evidence of evaluation being undertaken. It has been done by consent up until this point, and I think the thing that we need to do to strengthen the work of the IPA is to make it a mandatory requirement. At the moment that is not a mandatory



## HOUSE OF COMMONS

requirement, and I would very much welcome any recommendation that suggested that we do that.

The other thing is the MPRG and the TAP panels. The MPRG is chaired by myself and Alex Chisholm. As we sign off major Government programmes, one of our questions is always, "How do you know if it's going to work? What's your evaluation strategy? Demonstrate why you've decided, through the Green Book evaluation and your options appraisal, that this is the right thing to do." That is almost the first question that we ask at every single MPRG. We are currently updating the advice for the people who chair TAP panels as well. These are Treasury approval points that release funding for other projects that do not make the GMPP. So there is more work that we need to do.

On your very specific example, I am afraid I do not know why, but I would be very happy to go and have a look and provide further detail, if that is useful. But that is how it should operate.

**Q23 Sarah Olney:** Okay. We have other examples of evaluations that have not been carried out.

**Cat Little:** I am very happy to follow up on specifics.

**Q24 Sarah Olney:** It is obviously a serious issue. We are seeing Government spending decisions—and decisions to cut spending—being made and evaluations not being carried out.

More generally, Sir Tom, what have you been doing to crack down on those programmes that we have identified through the course of our work where a proper evaluation has not been carried out? What are you doing to ensure that that is not going to happen in the future?

**Sir Tom Scholar:** I would go back to what I said earlier about the spending review, and putting not just evaluation but a focus on outcomes—you cannot focus on outcomes without doing a proper evaluation—right at the centre of how we approached the last spending review. But that is not just a one-off exercise when deciding to allocate money; it is now part of the centre of Government's continuous monitoring of departmental performance.

There are lots of elements to that. One is the evaluation work that Catherine was talking about through the evaluation taskforce, but another is the requirement on Departments to publish their outcome delivery plans, which are deliberately focused not just on a spending line-by-line analysis of what they are doing, but on what outcomes they are trying to achieve in the world.

The world is complicated and many things are affected by the actions of many players, some in Government and some outside Government, so it is not in any way an easy thing to do. But our strong hope is that these various reforms that I have mentioned will improve the Treasury's and the centre of Government's ability to hold Departments to account for achieving those outcomes, but also improve your ability—that of Parliament and the public—to do the same. That is the purpose.



**Q25 Sarah Olney:** You were talking there about line items of spend, or whatever, but the example I cited on the courts and tribunals was where you had stopped spending. When you are looking at your evaluation strategy, is your evaluation approach as robust in terms of evaluating where you have made a decision to stop spending, and the impact that that is having, as it is on where you are spending, and the outcomes you are seeing from that spending?

**Sir Tom Scholar:** That is a very good question. The first answer is that when you are setting spending plans you are deciding both what to continue or increase and what not to continue or to decrease, but whether it is symmetric—Cat, perhaps you would like to comment.

**Cat Little:** The simple answer is yes it should be, but quite often there are gaps, as the Report highlights. Catherine might want to give a couple of examples.

Critically, evaluation has very many purposes, but one of its most important purposes is in the prioritisation of spend. Certainly at the SR we used evaluation; in fact we integrated the work of the evaluation taskforce into our assessment of spending bids. That allowed us to use evaluation, where it existed, to understand whether we should scale up or stop evaluation, and where there are gaps we are now following up as part of the SR conditions. For savings and efficiencies and cuts to services—changes to services—we fully expect evaluation to be part of the total project and programme life cycle. The big issue for us—this goes back to the primary question, really—is how do we address the gaps, how do we embed it within the culture of Departments, how do we ensure that it is followed through? Catherine might want to answer with some specifics.

**Catherine Hutchinson:** To follow up on the first question—how do we know whether Departments are evaluating effectively or not?—the evaluation taskforce has an account management model, so there are individuals within the evaluation taskforce who have regular engagement so that we can pick up any major priorities or programmes that are occurring in the Department and we can offer advice at the earliest possible point. That helps us build a relationship and intervene before anything eventually happens without good-quality evidence.

We also have several processes—Tom mentioned the outcome delivery plans. Evaluations are now being embedded into those reporting mechanisms and there will be published evaluation summaries for each of those outcome delivery plans for Departments. So that is another useful mechanism, each time, to make sure there is a culture of evaluation at the very earliest point.

For me it is a bugbear that evaluation is seen as a sort of navel-gazing exercise after something has happened, when actually the magenta handbook sets out that evaluation should be considered before you do something; while you are doing it, to see whether the impact that you intend to have is actually happening; and also afterwards, to gain any learning from value-for-money assessments. So, absolutely, the



## HOUSE OF COMMONS

evaluation taskforce is giving advice at that very early point to each Department that they have seen.

That includes, for example, the National Citizen Service, where we gave advice during the spending review about the quality, evaluating the impact, which led to a scaling down of that programme, and for us to then enforce a set of conditions that insist that they evaluate more robustly that service, to ensure that any further spending decisions can be made with the most important information available.

We have, I think, 15 different programmes and projects with certain conditions that we are asking in very specific areas. Particularly, the evaluation taskforce has looked at the top 10 priority areas where they think that evaluation is the most important to do, and we will be focusing specifically on those areas, which will enable us to be much more proactive with Departments in areas that could make the most significant difference.

**Professor Sir Ian Diamond:** Could I just say that Catherine made an incredibly important point there? That is that evaluation isn't something that is just done afterwards; it absolutely has to be part of the entire process. I always say that the critical thing is clarity of thought, because if you haven't got a clear view when you start of what success looks like, which is in a very broad sense, there is little chance that your programme will work well. That is a very important part of evaluation—and also demanding that it is not just about looking at inputs; it is about looking at outcomes and impacts. So Catherine's point is so incredibly important.

Q26 **Sarah Olney:** Sir Tom, we have talked about how evaluation is part of settlement conditions at the spending review. What do you do if those conditions are not met? What would Treasury do? Would you not consider future bids? How would they pay the price for that?

**Sir Tom Scholar:** We have several tools. The ultimate stopgap is that where funding has been allocated in a conditional way—it always is conditional, as set out in the settlement letter—and where one of the conditions is proper evaluation, and that evaluation is not carried out, then ultimately that spending is irregular, and then that becomes an issue for the NAO to look at when they are signing off the accounts. That is the ultimate backstop.

We obviously hope to sort things out before we get there. There is ongoing monitoring by the spending teams. As Cat said, the spending teams have their list of conditions that they need to see fulfilled—they look after that. Cat might want to talk about the regular meetings with finance directors. We do an annual assessment of the accounting officer's performance against all professional standards. I think the analytical function produce a report as part of that; the finance function produce a report as part of that. It is all part of the accounting officer's appraisal, ultimately. So there are a whole series of tools, and we have the ability to escalate.

Q27 **Sarah Olney:** So there would be several different people having a view



## HOUSE OF COMMONS

on whether the evaluation conditions have been met—the accounting officer but also the functional lead.

**Sir Tom Scholar:** The three key people are the accounting officer, the finance director and the head of analysis. Depending on precisely what the nature of the problem is, you would look to one or other of those three people to take the lead in solving it, but ultimately it all goes to the accounting officer who is accountable for the performance of the Department.

**Cat Little:** The two biggest levers that hurt Departments the most that we can take if there is non-compliance or if we think standards haven't been met are the obvious one of stopping funding—it is not unheard of for us to reduce funding or to make changes to the funding settlement—and the delegated authority limit that we set for each Department. As this Committee knows, that is applied in different ways to different types of expenditure. We are undertaking the most comprehensive review of delegated authority limits at the moment for this forthcoming financial year, which will take into account compliance with conditions, breaches, any irregularities. Ultimately, it is an earned autonomy regime, and the more you comply, the more autonomy you have. We think that is a very fair and open and transparent way of doing it.

Q28 **Chair:** Can I ask what has driven that review?

**Cat Little:** We hadn't done it for three or four years, and we felt it was timely, given everything—

Q29 **Chair:** We were looking in detail at the Department of Health and Social Care accounts this week and there were areas where compounded spending breached the limit and the Treasury took a dim view of some procedures that weren't followed, for smaller contracts, which cumulatively added up to the limit. Has that been a factor in your thinking?

**Cat Little:** It is—yes. The Department of Health is obviously a Department that has a range of different issues that we will take into account as to how much authority it has going forward.

Q30 **Chair:** So you might be clipping the wings of the Department of Health.

**Cat Little:** We may do, and, equally, we may raise delegated authority limits for some Departments that have shown good practice and complied with the conditions that we set.

Q31 **Chair:** I don't know if the CAG wants to come in on that issue about the regularity of spending and review.

**Gareth Davies:** Obviously, the year that the Committee was looking at for that particular Department was an exceptional year, so I think you were making it clear that you wouldn't expect repetition of things that happened in an emergency situation in the future. It was an unusual concentration of issues in one set of accounts, but I think we covered that on Monday.



## HOUSE OF COMMONS

Q32 **Chair:** We will keep a very close eye on that, because it is quite interesting. I suspect there may be resistance from accounting officers.

**Cat Little:** Well, ultimately—

**Chair:** They can resist, but they can't resist. Sir Tom gives that knowing smile.

**Cat Little:** Obviously, we do this in a very open, rational, evidence-based way, but ultimately it is not for debate.

**Chair:** Okay—a strong Treasury today!

Q33 **Sarah Olney:** Quickly, on that point about evaluation standards, Catherine Hutchinson, do you have any input to that assessment of how well the evaluation standards set out have been achieved?

**Catherine Hutchinson:** There are two halves to that answer. The first is that we work very closely with Treasury colleagues. When Departments are giving their responses to those evaluation settlement conditions, they will go to the evaluation taskforce, which will then review to ensure that those conditions are being met properly so we can help them to be intelligent customers—to understand whether those evaluation plans are as robust as they can be.

In addition, you are absolutely right that the taskforce also has a role in enforcing standards within Departments itself, as there is much more to departmental activity than just the spending review activity. We have a responsibility. The magenta handbook is the guidance document for all quality and expectations for evaluations, so every time we receive any evaluation plans or strategies, we will also always ensure that they are complying with the handbook, which is owned centrally. We would also then give very clear support, through the ODP process, the SR process and our account manager model, to ensure that we are insisting—and, every time a departmental reporting mechanism happens, that we are checking—that that process is as good quality as it can be.

We have a publicly available website that will also instantly demonstrate whether Departments are being compliant, as we will publish all evaluation plans, strategies and trial protocols. That will enable the public to transparently, and in real time, see whether Departments are complying with those requirements.

Q34 **Chair:** So that is your power—publication.

**Catherine Hutchinson:** Hopefully, yes.

Q35 **Chair:** You have power inside the system, but if an accounting officer pushed back, how far can you push, or does it go through the other structures?

**Catherine Hutchinson:** If it was a spending review requirement, we would go through Treasury colleagues because, essentially, they have to comply, for the reasons that Cat and Tom have outlined. If it is not a spending review requirement—let's say it is in our top 10 and they don't



## HOUSE OF COMMONS

comply—it would be through our very visible publicly facing website, where we would outline our priority areas and the progress against those. Ultimately, we report to the Government modernisation and reform board, which is chaired by the Cabinet Secretary. He checks whether the evaluation taskforce is delivering against the expectations in that modernisation and reform agenda. We, again, have very good, clear senior support and senior visibility of compliance with the evaluation taskforce requests.

**Chair:** Okay, there is always a matrix.

**Sir Tom Scholar:** In the circumstances that Catherine is talking about, we in the Treasury would certainly be playing a full part in that too. We are a single team here, trying to do the same job, which is to drive up standards.

**Professor Sir Ian Diamond:** If we felt there was an evaluation that perhaps had not been undertaken to the highest standards, there is also the opportunity for the Office for Statistics Regulation to take a view on the quality of the work.

**Chair:** Which, of course, you have done—and robustly.

Q36 **Sarah Olney:** Just to come back to the modelling and who is responsible for it, we have the Aqua Book with the modelling guidance, but you are not using that when you are assessing spending review processes, or it is not part of the investment approval process. Is that correct? In terms of the guidance in the Aqua Book, you are not checking to see if that has been followed with the modelling.

**Cat Little:** Perhaps just thinking about how we actually use models in a lot of our work, we use financial models at fiscal events. Primarily, we rely on the assurance work that I described in the first and second lines—

**Sarah Olney:** The three lines of defence, yes.

**Cat Little:** The first thing that we would always do would be to say, “O Department, how have you gone about assuring your models?” and we would take a view as to how much reliance we would place on that. Obviously, as part of that conversation, we would want to be assured that they have taken the Aqua Book and its guidance into account, but we are not specifically checking. That comes back to the role of the accounting officer, the analysis function and the finance director, in ensuring that they are undertaking their personal responsibilities for the quality of the models they are using for decision making.

There are some examples, particularly at the Budget and the SR, where we undertake to produce costing estimates. That is where the OBR relies on our use of models—sometimes applied by Departments, and sometimes developed in the Treasury—and we have those costing notes. They are independently reviewed through an iterative process between the Treasury, the OBR and Departments, and they ultimately critique and assess those costing notes. Ultimately, we want to make sure, as part of



our year-end accounting officer and functional assessments, that the Aqua Book is being taken seriously and people are using it.

- Q37 **Peter Grant:** Sir Ian, could I come on to a very live and significant piece of modelling that had to be done a couple of years ago, which is all the modelling around the covid pandemic? In fact, it must have been almost exactly two years ago today. How important a role did statisticians and other professionals have in the modelling that led to decisions about which restrictions had to be brought in and which did not?

**Professor Sir Ian Diamond:** It is worth saying that the models that you describe were put together by a sub-committee of the SAGE committee, which was called SPI-M. This country has historically invested, largely through the Medical Research Council and the National Institute for Health Research, in groups in different universities that have been extremely good—world-class, I would say—at modelling epidemics. Other countries—for example, Australia—had to bring in people who had been modelling other phenomena, but we had a number of groups that were absolutely world-class in modelling epidemics and used different approaches. Some would take some data and build equation-based models, which they would then project forwards. Others almost set up the world like a computer game in some ways, to see what happens if people mingle and mix.

When the sub-committee of SAGE was put together, it took the different groups—for example, groups from Imperial College, the London School of Hygiene and Tropical Medicine, Warwick, Edinburgh and Lancaster—and met weekly in order to come to a consensus. As the minutes of SAGE will show, the first item on every SAGE agenda was a report from Professor Graham Medley, who was the chair of the sub-committee, on what the consensus was. That consensus was a matter of different assumptions, slightly different datasets and different approaches, so it was effectively getting its variation from using different approaches to address the same question, using slightly different datasets. I believe that that brought with it independence, and it brought with it the benefits of the UK having invested, over a long time, in that world-class skill. It also brought the fact that we were not just saying, “We’ve got one person over there who does this.” We were actually saying, “We have different teams using slightly different approaches, and we will come to a consensus around what the range of outcomes could well be.”

- Q38 **Peter Grant:** Was the same process used later, when consideration was being given to lifting to some the restrictions? Would there have been teams who would have modelled what happens if we reduce social distancing from 2 metres to 1 metre, for example?

**Professor Sir Ian Diamond:** Exactly so. You do not always have the data you need in order to impact but, as I indicated, the minutes of every SAGE meeting will show that the first item on the agenda, more or less, was this discussion. The discussions at different times were about how we stop the increase in the pandemic, and at other times, given that we were clearly on a reduction, they were about what the impact would be of, for example, reducing social distancing or allowing the schools to go back.



What would be the impact of that happening? The models would come out with different estimates of what might happen, and a consensus would be reached.

- Q39 **Peter Grant:** As well as the modelling of the behaviour of the virus and its likely spread through the population, there was also a lot of modelling done on financial and economic matters. For example, asking, “If we take the decision to shut down a lot of businesses, what will the economic impact be given time?” I appreciate we have Sir Tom Scholar here, whose Department did quite a lot of that work. Would there have been a similar process, with the same degree of rigour, applied to that modelling?

**Professor Sir Ian Diamond:** I am going to pass that across to Sir Tom, because all the economic modelling was done by the Treasury. I will say that the Office for National Statistics had a very good and positive relationship with the Treasury throughout the pandemic about what economic data was needed. For example, very early on in the pandemic, the Office for National Statistics did a business impacts of coronavirus survey, which was a large survey that went out to a large sample of businesses, fortnightly. The Treasury, along with other Departments, was heavily involved in the design of the questionnaire to make sure that the data that we needed was impacted into those questions. In another example, we were able to access a very large proportion—over 90%—of the credit and debit card transactions in the country, to understand how the consumer was responding to different aspects of the pandemic. That was all done in an ethically sound and privacy-enhanced way. That data was then able to go into the Treasury to impact and help the modelling that was going on. Having said that, I will pass over to Sir Tom.

**Sir Tom Scholar:** I wrote a letter on this last year that set out in some detail how we approached the task. I will give you the short version now— if you are interested in the long version, I will happily send the Committee a copy of the letter. The short version is that the Treasury was involved throughout in advising Ministers on the potential economic impacts of the various things being considered. That advice incorporated, to the best of our ability, economic analysis and, in doing that, we drew on a lot of the work that Ian has talked about, and we also talked a great deal to the Bank of England and the OBR. We surveyed all of the work being done in universities and financial institutions.

It was a situation that was exceptionally difficult to model to any satisfactory degree, because in any economic model you are trying to make predictions for the future based on your experience of the past. There was no experience, no set of reliable estimates and no previous models available for that situation. All of the estimates were subject to a very wide degree of uncertainty. What we were able to do, whenever a measure was considered for a particular sector or region—or a particular sector in a particular region—was draw out the basic information on the likely direction of effects, the importance of the sector in that region and the possible knock-on effects elsewhere. We were able to develop ranges of impact.



## HOUSE OF COMMONS

The task of trying to estimate the effect on the economy is rather different from the epidemiology itself, for which there is a huge wealth of experience in this country and many others. There are obviously still uncertainties, as we all saw over the last two years. The nature of the exercise is quite different. If it is helpful, Chair, I can send you a copy of the letter that explains all that.

**Chair:** That would be very helpful. Thank you, Tom.

Q40 **Peter Grant:** As a statistician, Sir Ian, looking at both the public health modelling of this deadly virus and the modelling done on economic and other factors, was there anything that didn't work well that we need to learn from for future modelling?

**Professor Sir Ian Diamond:** An enormous amount of extremely good modelling was done. For me, it is critical to get as much data as quickly as possible. As Tom has just outlined, one is dealing with a new situation. We've dealt with a new situation here three or four times in the epidemiological world, because you think you're getting on top of something and then you get a new variant. For example, when you get a new variant, we could very quickly identify that variant through the ONS coronavirus infection survey. However, it will then be a little while before the virologists are able to say how challenging it will be and the extent to which there will be hospitalisation.

I have heard one very famous epidemiologist say that we have to acknowledge that omicron has cut us a break. There was no reason why omicron should have been so mild—I hesitate to say that, for those people, including myself, who have had it and been pretty ill, but it did not lead to such hospitalisation and mortality. We didn't know that initially because it started in South Africa, and there was some thought that there hadn't been so much mortality because of the different demographics there.

My point is that I do think we have done well; Tom's teams have also done well. However, in future, we need to really be agile about data. We have been agile, but we need to be ever more agile. We need to be faster at sharing data and make sure that we have independent scrutiny of our models and statistics.

Q41 **Peter Grant:** We had the permanent secretary of the Department of Health and Social Care before the Committee a few days ago. He was almost frighteningly frank about the fact that, if we'd been hit by a variant that was as transmissible as omicron and as lethal as delta, we'd all be in a very different place.

**Professor Sir Ian Diamond:** There is absolutely no question about that. I am not a virologist, but, from talking to world-leading virologists in order to inform our own modelling, my understanding is that when omicron came along, until we actually did some work, it could have been more, equal or less—and the same with the next variant.

Q42 **Peter Grant:** When we look back to the response of Departments right at



the start of the pandemic—which was one of the things that we gave Sir Tom and others quite a hard time about, at the time—almost all of the economic response and the financial support packages and so on had to be done very quickly, because nobody had been prepared for a virus such as covid. A lot of work had been done in preparation for an outbreak of a flu-type virus; the main difference is that if you’ve got flu, you know there’s something wrong with you, so you’re less likely to go out and about and infect other people. It’s very possible to be highly infectious with covid and not know there’s anything wrong with you—asymptomatic transmission is how they describe it. These plans had been put together five or 10 years in the past, and it doesn’t look as if anybody had recognised the possibility that when the pandemic came, we might meet a virus where asymptomatic transmission was so prevalent.

Clearly, we have learned that lesson about planning for a future pandemic, but how do we ensure that in any modelling or forecasting, which is so critical—literally life and death for some people—that all the bases, all the likely scenarios and all the unlikely ones have been identified, rather than simply preparing for one scenario and having a blind spot?

**Professor Sir Ian Diamond:** I absolutely do not know how we are going to get to all the unknown unknowns, so I would not claim that we could now plan brilliantly and perfectly for any outcome in the future. Having said that, there are a number of things we have learned, which I am sure Tom will be happy to add to.

For example, in one week in April 2020, we set up a community survey—the only country in the world to have a survey—which enables us to properly understand the asymptomatic nature, which enables us to say right across the UK what the level of positivity is. That was set up from scratch. At the moment, we will continue to do it for one more year. I believe that as part of planning for the future, one could have, for example, a public health survey, which is not expensive, that is ongoing and ready to surge whenever you need it. I think there are a number of things like that, which when we look very carefully at lessons learned we would want to think, “We can do that.”

I am sure there will be unknown unknowns, but at the same time I do think there are a number of areas that we could say now, “We need to move forward.” Part of that is going to be better data sharing and the better ability to data share. In the spending review, the integrated data service, which the Treasury very generously funded, is an important area for the future, which will enable us to attract things and respond quickly.

**Peter Grant:** I will leave it there, Chair. I may want to come back to Sir Tom later on.

Q43 **Kate Green:** I would like to explore a little more the way in which you are embedding evaluation practice and culture across Government. Could we pick up on what you were saying a few moments ago about evaluation being important before, during and after the delivery of a programme?



## HOUSE OF COMMONS

First, why do you think there is a perceived reluctance identified in the NAO Report in some quarters to publish the findings of evaluations when they have been conducted? If you agree that that is the case, what needs to happen to ensure there is transparency for the public when evaluations are undertaken?

**Professor Sir Ian Diamond:** I absolutely believe that transparency is critical. It is critical for three reasons. First, if things go wrong but the project has been professionally run, we learn an enormous amount. We do not have the monopoly on knowledge that every project is going to be brilliant. Sometimes we have taken a view and we move it forward, or maybe it did not work in some ways or in the way we expected. Those are lessons learned and they should be available for the future, so we do not make the same mistakes again. Secondly, we are—everyone here—the guardians of the public purse. It aids trust from all our fellow citizens in the way we look after the public purse if we are transparent about how we have spent it and where things can be evaluated. Thirdly, I think it is incredibly important that we are transparent in every way, because we should be open to challenge. I believe that that is undertaken both by having the protocols available and the data for secondary analysis, as well as ensuring that one has independent peer review of any publications.

**Cat Little:** I am sure Catherine will want to talk about some specifics. The really important thing for me, and the NAO Report helpfully sets out the supply and the demand barriers, is that you have to work quite hard at both sides of the equation consistently. Quite frankly, while there has been brilliant momentum in the last couple of years, it is not until recently that we have had lots of levers, lots of follow-up, and lots of political engagement on why evaluations had not been published. Tom referred earlier to the letter that the Chief Secretary and the Chancellor of the Duchy of Lancaster sent before the SR. Specifically in that letter they asked Departments to set out the schedule of evaluations that they were due to publish and, if they were not planning to publish them, whether they could please explain why and what the barriers were.

That is the first time that I am aware of the centre and Ministers asking those questions. We are currently going through all those responses and looking at how we can help. Your point about culture is absolutely critical. Again, it has only been in the last couple of years that the functions have been very hard in driving through upskilling evaluation practice and embedding it into a lot of what we do. We need to have consistent momentum over a period of time to make a difference, but Catherine has been doing some specific work.

**Catherine Hutchinson:** Essentially, you had two questions. One was about why there are not more evaluations before, during and after. The other was what we are doing about transparency. To your first question, there are a lot of perceived barriers that the evaluation taskforce has been working to try to overcome. I mentioned the top 10 evaluations—the ones that we thought were the most important for us to get behind. We specifically chose at least a few of those to tackle those perceived barriers.



## HOUSE OF COMMONS

Two of them are going to be chosen on the basis that they are cross-departmental or involve data sharing. As Ian rightly said, that has been a barrier before, but there has been significant investment and advancement in those areas, so we are trying to push Departments, essentially, to work in those areas, exploit those opportunities, and show that that is not a barrier anymore. Secondly, we are going to have two that are evaluations in areas that are traditionally very difficult, or perceived to be difficult, to evaluate. Frequently, people do not evaluate because they say, "Oh, you can't evaluate that; it's too complicated," or, "You can't evaluate that; it's a particular case." We are trying to target areas where they try to put those sorts of evaluations into those buckets. That is the first practical step.

Q44 **Chair:** Can you give us examples? Are you allowed to tell us? We are asking you. Tell us, please.

**Catherine Hutchinson:** We are awaiting ministerial sign-off for some of those, so as long as you understand that they may not be the same, an example of a hard-to-evaluate area would be the future of farming, because it is about environmental and social impacts. Some of those things might be intangibles, but for us they are not intangibles. A good example of cross-departmental data sharing would either be in children's social care or in reducing reoffending, which makes the previous question about courts closing quite an interesting one for us to take back.

On the second question about transparency, I mentioned the website. That is going to be really useful, as Ian mentioned. Having protocols published creates a level of transparency, so when you are saying which evaluations you are starting there is almost no way of not publishing once you have already published that they are starting, because people are aware of what is happening. The issue is more that evaluations have happened that we do not know about and that we cannot get to because they have never been published.

For us, it is important to have that transparency right at the very beginning, because everyone thinks that their project is going to be successful. That is why they have chosen to evaluate, so they are quite happy to publish that they are looking into it at the beginning. It is only later, at the end, maybe when there is a change, that they might change their mind, so for us it is very important to get a clear process in place. It is about having the models, engaging the Departments, and reassuring them that being able to publish and to be open and transparent is the best way. For us, the Government have been more transparent than ever before in taking those steps.

Q45 **Kate Green:** You talked a bit about data sharing, but in terms of sharing good practice, structuring evaluations, gathering data and assessing which data will be relevant, what work is being done to support a cross-Government approach to raising standards?

**Professor Sir Ian Diamond:** The national data strategy, which is owned by the Department for Digital, Culture, Media and Sport, has been

excellent in having 10 facets to it. That is important because part of it is actually data sharing, but if you want to share data you need data standards. Let me just be trivial: if you register your date of birth in a different way from how someone else registers their date of birth and we then try to join them up, we have to go through an enormous process. Having a real set of data standards across government will be incredibly important, and the Data Standards Authority is making really good progress. On top of that you need privacy and ethical standards.

All this can be done and needs to be done in a way that does not hold things up. I think the Government is doing very well, led largely within the Cabinet Office and DCMS to take the national data strategy forward. As part of that, we at the ONS lead the integrated data service, which is the technology of bringing bits of data together. For example, the analysts looking at this multidisciplinary problem will be able to get data from different parts of government, bring them together, link them, and then be able, as in the reoffending example, to do unbelievable analyses of the sort that I could only have dreamt about 20 years ago. I think we are making very good progress, but the support of the national data strategy will be super-important.

**Q46 Kate Green:** Are you confident that within government the skills and the potential to develop the skills of the workforce in the future are properly in place to manage and share data and learning across Departments?

**Professor Sir Ian Diamond:** I start by saying this is not just a problem for government. People with digital data and statistical skills are in massive demand everywhere. There is the analysis function, statistics, economics and things like that as well as the digital and data function, which is led by Joanna Davinson in the Cabinet Office. They are doing a great job of thinking through the whole capability—not just thinking through, not just saying, “We’re thinking about it,” but taking actual actions and really working on apprenticeships with FE colleges. We at ONS have a good one with Cardiff Metropolitan University that is bringing people in, and that is common across other Departments. There is a real joined-up strategy for increasing the number of people with those skills at all levels in government.

That is only so far, though. We need, as I indicated earlier, to have continued professional development and a culture where you expect to move forward. That goes not only to what evaluation is and how we do evaluation, which is a really important one where we need to upskill many people, but also do some of the statistical analysis techniques. That is where the analysis function comes in. Also, the finance modelling that we talked about is where the finance function comes in.

**Catherine Hutchinson:** We have had significant support to create a new culture that you need for evaluation. You do not just need analysts; you need all the other professions on board. The evaluation taskforce recently ran a cross-government conference, a policy that works conference, and over 2,200 civil servants signed up to that four-day event, which was a phenomenal response. We had the Chief Secretary to the Treasury, the



## HOUSE OF COMMONS

Minister for Brexit Opportunities and Government Efficiency, and the Secretary of State for DLHUC all presenting, as well as four permanent secretaries. There was significant resourcing and ministerial backing to get to this moving in a way that I don't think there has been previously, so I think we can be much more confident than we have ever been that this is the case.

- Q47 **Kate Green:** On your point, Sir Ian, about standardising the way in which data are captured, of course we are dealing with some very long-tail datasets that are really important for us in, for example, tracking cohorts or changes over time. Where that older data might not be held in a format that is matchable with the way data gathering has taken place subsequently, how do you deal with issues like that where you have breaks in datasets or inconsistencies?

**Professor Sir Ian Diamond:** I used to work a lot on cohort data. If I had a PhD student who came with a new wave of a cohort, we would say, "It is going to take you 18 months before you can do anything much, because you're going to be cleaning the data." It is a real challenge. I think we are getting better at it, because technology has enabled things to improve, but often there is absolutely no alternative to painstaking data engineering. Sometimes, actually, you just have to sit there with the data and change the numbers.

- Q48 **Kate Green:** Is there the will to commit the necessary resource to do that? Do people see the importance of being able to track data back?

**Professor Sir Ian Diamond:** The answer is yes. I do think, sometimes, we don't realise that in this country we have global jewels in a number of national cohorts—cohort '46, cohort '58, cohort '70, cohort '90 and millennium. We have real skills at doing this and we use those data a lot.

Now, what you often find is someone who has their own little cohort that they have been collecting; then, there has to be a judgment as to whether those data are really going to add enormous value. Are we linking everything? No, we're not. But do we have the ability to link where we believe there is value from so doing? 100%. Can we look backwards and link data? Yes. For example, we have an ONS longitudinal study that links census data from '71, '81, '91, 2001, 2011, plus vital registration data and council registrations and all sorts of things like that. At the moment, that is based on just four secret birth dates, but my intention, because I want to build inclusion into it, is to expand that to include a much wider range of the population. At the moment, around one in 100 people are in it. We would like it to be a much bigger proportion and we have the ability to do that. I am not quite sure how many censuses we could go back. We recently published a digital 1921 census.

So the answer is yes, where we think there is real value from doing it. It is painstaking, it is hard work, but the will is there.

- Q49 **Kate Green:** It is welcome to hear about the practices that have been developed and the changes in culture that are being fostered. To go back to the issue of publication and transparency, if a Department is carrying



## HOUSE OF COMMONS

out good evaluation, has specified it well, has done it well, has learned from it, but is none the less reluctant to publish it—we saw that recently, for example, with DWP, which was quite reluctant to publish some insights it had into disabled people’s attitudes to their treatment—what sanctions or powers do any of you have to make sure that evaluation is made public?

**Cat Little:** It really does depend on why it hasn’t been published. The sad fact is that ultimately a lot of this is down to ministerial consent—sorry, that is not the sad bit of it; that is the fact. We would like to see evaluation as analogous to statistics, but one is governed by a regulatory position, and one is entirely by mutual consent. That is the way in which the framework operates. Of course, we want to have dialogue with officials, and politician to politician if needed, to try to understand why documents and evaluations are not being published, but that is how our system works.

Q50 **Kate Green:** Given all the efforts and improvements that you are making, the establishment of the evaluation task force and the work that you, Catherine, and colleagues are doing, what is good going to look like and how will we know when you have brought about the improvements that are needed and that you are seeking to drive?

**Catherine Hutchinson:** That is an excellent question. I probably will do myself out of a job, but essentially, the ultimate goal will be that the ETF does not need to exist—that you should not have to have a central body that checks, pushes and works relentlessly to make a culture change, because that culture change should have happened. There are already excellent evaluation teams within Departments who are doing much of this work, but that needs to be more consistent. We have published on our website a clear theory of change about all the inputs and activities that the ETF is doing and what outputs and outcomes we hope will happen as a result. Some of those metrics might be about the number of programmes we have given advice to—up until the spending review we had given advice to 151 different programmes covering £35 billion-worth of spend—but how much is good enough is not really the question; it is more about the culture change and essentially the fact that the ETF should not be needed in the future, because this should be happening. We should be encouraging all the systems, processes and practices to be embedding evaluation into them so that it becomes second nature, a habit, rather than being something that is new or slightly irritating.

**Cat Little:** The interesting thing about the evaluation task force is that we will evaluate it. I am chair of the oversight group that oversees the evaluation task force and the Government reform ambition. We have a suite of KPIs that we are using to assess the task force progress and we will publish, ahead of the next SR, the evaluation of the evaluation task force.

**Catherine Hutchinson:** There is some other stuff that we are doing. For example, the generous £15 million that the Treasury have given the task force should be creating actionable evidence. Very soon, you should very



## HOUSE OF COMMONS

easily be able to see some of the impacts that we are having on specific areas.

Q51 **Chair:** Before I go to Mr Grant, I want to ask this question. You have got your money so far, Ms Hutchinson. How long is the task force envisaged to last?

**Catherine Hutchinson:** That's a good question.

**Chair:** Obviously, this is an opportunity for you to bid for a long career as head of the task force.

**Catherine Hutchinson:** Exactly!

As far as I understand it, the task force has been funded in this spending review round, so we have at least another three years. We have our metrics, our key performance indicators, where I hope we will be able to demonstrate our impact. Then, ideally, we would continue at least for a little while longer—I don't know how quick the culture change will be, but I doubt it will happen in three years—but as I say, I hope that we do not exist indefinitely.

**Chair:** Well, "task force", by its nature, sounds urgent and short term, but yes, by the sound of it we are talking about probably another spending round at least. You are right: the culture is a challenge. Mr Grant, over to you.

Q52 **Peter Grant:** Sir Tom, you have answered some questions from Kate Green about transparency. We are trying to get to the bottom of where accountability actually lies if a Department does not appear to be following the requirements. In relation to published registers of business-critical models, the NAO found that only nine out of 17 Government Departments have published them at all, and of the nine, five have not been updated for at least five years. First, is the Treasury one of the four that have an up-to-date business-critical model register published?

**Sir Tom Scholar:** No, we are not up to date. We published our list in, I think, 2013, at the time when the requirement was first introduced, and—I looked into this just the other day in preparation for this Committee—we have not updated it since then. I think that is a fault of omission—in other words, people just forgot. I don't think there was any particular reason for it, but we will put that right and we will publish our list in the next few months.

Q53 **Peter Grant:** It doesn't look too clever if the Treasury is one of the Departments that are not following Treasury advice on managing public money, does it?

**Sir Tom Scholar:** We should have updated it. We haven't and we will put that right.

Q54 **Peter Grant:** The NAO also looked at a sample of 75 business-critical models, and for 45 of them—more than half—they could not find any publicly available information on the models. Does that suggest that



## HOUSE OF COMMONS

whatever arrangements for accountability, and enforcement if need be, are in place just now are quite clearly not working?

**Sir Tom Scholar:** Well, the whole thing we are discussing this afternoon is trying to improve efforts in this regard. I think, again, we should look at what the nature of the models is. To take the Treasury as an example, we have 66 business-critical models, of which slightly under half are operated by our finance and HR departments in order to assist the running of the Department. In a sense, the output from those models is the functioning of the Department and the accounts that the finance team produce and the NAO audits. That is obviously a very different type of model from something like the model to calculate the Barnett consequentials of spending in England. The outputs of that are published, and there is also a lot of information available publicly about the model, how it operates and so on.

A business-critical model can cover a very wide range of different things. In some cases, there is a clear public interest in information about it being available. Often there will be a public interest in publishing the model so that other people can use it. The OBR do with their model of the economy, so that other people can use the OBR model to develop their own results. If it is something to do with advertising jobs internally or predicting staff flows internally, that is a little bit different. But I think what we need to do as we are looking at this, following the NAO Report—obviously, we will get the PAC Report as well—is work out what the guidance is to Departments on the models that really do require having greater transparency around them, and then push for that.

Q55 **Peter Grant:** The Government's view is that transparency should be the default position—look, for example, at the way that the freedom of information legislation is set up. For how many of the models that you have within the Treasury could it be argued that there is a possible detriment to public interest in publishing? Are there any that it would be against the public interest to publish?

**Sir Tom Scholar:** I can't immediately think of why that would be the case, but I would need to go away and look at it. In some cases, I would also need to understand what precisely is meant by "publishing". The model might be a very simple thing, but it depends on data. Obviously, some of it is public, and some of it is not. Sometimes there are good reasons why data is not public.

Q56 **Peter Grant:** Finally, I was struck by the figure on page 31 of the NAO Report, where they asked the chief analysts and heads of policy profession in all 16 core Government Departments what they saw as being the barriers to using evaluation as part of decision making. I know it is a relatively small sample, but there are striking differences in the reasons given by members of those two professional groups. Interestingly, the people who needed to use the analysis said one of the main reasons was that the analysis often was not very good. Not surprisingly, that was not an answer that was given by the people who did the analysis. How big an issue is it if, even between colleagues



working together in the same Department, there appears to be a significant difference of opinion as to what the barriers are and, therefore, about what needs to be done, and who needs to do it, in order to get where we want to be?

**Sir Tom Scholar:** What we are talking about here is improving the ability of policy makers and analysts to work together to improve the joint fruits of their work. There are some barriers to that, as the Report says. If you ask those two groups of people what they think is the main issue—well, it is human nature that you tend not to say, “It’s basically my fault.” You tend to look for other reasons, so that is not completely surprising.

Ian might want to talk about this as well. What we are trying to do, both through our work here and through the work of other people in the policy profession, is to improve the ability of each group of people to understand the work of the other so that each is a more intelligent customer of the work of the other and, ultimately, they work better together as a single team.

To give a historical example, when I started work at the Treasury in the early ‘90s, then, as today, we had policy people and analysts, particularly microeconomists, working on public spending, but they worked in different teams, so we had the transport spending team and, in an entirely separate part of the Department, we had transport economists. That is no longer the case. We now have a single transport spending team that includes economists, statisticians and policy analysts all in one team, and that is a much more effective way of doing things.

Ian, in the analytical profession, is trying to improve analysts’ understanding of the policy side, as he mentioned earlier. Tamara Finkelstein, running the policy profession, is stressing the importance of proper analysis and evaluation as a collective, across-the-board effort.

I would just add that the UK Government are not unique in experiencing this. The Report refers to an OECD study that has pretty similar findings from other countries, and my guess is that you get criticism of findings from other businesses, not just Government but other people who have the same issue of trying to bring together policy and analysis. It is an issue, but we need to keep working on it.

**Professor Sir Ian Diamond:** I agree with everything that Tom has just said and I never, ever criticised the data from a potentially small sample, but I might add that in the functional survey, which was quite a large survey, the analysis function gets the highest rating from other people, and in fact lower from the analysis function itself. It is appreciated, but as Tom has said a couple of times today, it is getting people working around the same table from day zero, so there isn’t someone being dragged in to help; they are all in a team, getting the chemistry right, and then we will get better policy.

Q57 **Peter Grant:** Finally, Sir Tom, one of the possible barriers that the NAO asked these professionals about was that the evaluation evidence doesn’t



## HOUSE OF COMMONS

support current policy or political direction or commitments. A minority, but still almost one in four, of the professionals who took part in the survey thought that that was an issue. How concerned are you, and how concerned should we be, if a significant minority of senior civil servants think that part of the reason we are not publishing more evaluation plans at the beginning is because the politicians think that the evidence doesn't support the decision they've taken?

**Sir Tom Scholar:** First, as we have been at pains to stress this afternoon, we are all fully behind greater transparency and pushing for that, and I think we are making progress. Secondly, that is also an explicit aim of the Government as a whole, including statements from senior Ministers on Government reform and the completely central role of evidence-based policy making, proper analysis and evaluation.

It would not surprise me at all that there are particular examples of Departments that find it inconvenient to publish inconvenient analysis; that's life. The Government as a whole are trying to correct that by establishing default positions, and we are the people who are largely instrumental in that. Parliamentary Committees and the NAO also have a really important part to play in that. It is reasonable to know that that is the perception, but I think it underlines the need to push further.

Q58 **Chair:** Professor Sir Ian, can you explain to us in layman's terms how you would evaluate whether modelling improvements are actually really improvements? If we are trying to get better outcomes, you have got to evaluate your own modelling, but who is judging that and how are you going through that process?

**Professor Sir Ian Diamond:** At the end of the day, I would start by looking backwards. In other words, we have a model, and we have historical data, which means we know what has happened, so let's go back and see how different models do. That is incredibly important, because you need to have confidence when making a forecast that you have some slip. I always start by calibrating against the past, saying, "Does my model work for the past data?"

Q59 **Chair:** So you would literally run those models through past data?

**Professor Sir Ian Diamond:** 100%. When you are confident that you are getting to where you would expect to be and are making good analyses, it is always good to see if somebody else is doing this. As I mentioned earlier in response to Mr Grant, the modellers use slightly different techniques, but they do compare. I would be disappointed if I was modelling something, which I used to do a lot with different statisticians, and we did not say, "Let's look at the same data in a slightly different way and see if we get the same kinds of answers." That is how I would build confidence before moving forward.

Q60 **Chair:** There has been criticism from certain quarters at times of the covid modelling that SPI-M did. Do you think that is fair? Do you think in retrospect there are things that could have been done that would have built confidence in modelling in the sceptical group?



**Professor Sir Ian Diamond:** I think there's always an opportunity to sit down and have a conversation with people about modelling to properly understand what the criticisms are. Equally, I do think that the pandemic has been a situation where, unlike myself, many people in the population have hindsight. I fear that, as ever, there are people on different sides of the fence, and you cannot do anything more than to be transparent. You have got to be transparent and say, "What are the assumptions?"

At the end of the day, a mathematical model depends on data and assumptions. You and I may disagree on two assumptions. I would then suggest we run the model with both assumptions. We would then see what the difference is and be transparent about it. Ultimately, there is not an algorithm I know that tells you which of those two assumptions about the future is right.

Q61 **Chair:** But it gives you the range of actions?

**Professor Sir Ian Diamond:** Exactly right. Ultimately, it is a judgment between you and me. I might say that if we open schools, we will not see too much more mixing between generations in secondary schools, but we might in primary schools. You may disagree. Ultimately, I would run both assumptions and see what the difference was. I cannot tell you which of those assumptions is going to be right until the world has happened.

Q62 **Chair:** Thank you; that was really helpfully explained. I want to ask you, Professor, about funding for the analytical function. Do you think you have enough resource and power to help drive this culture change through Government?

**Professor Sir Ian Diamond:** I don't, I'll be honest. However, there is good news. Initially, the professions were so keen to push forward the analytic function that under my predecessor the Office for National Statistics said, "Look, this is so important. We will find a small amount of money to have a small secretariat to drive it forward." That is where we are. The agreement we have is that, following the spending review and following some of the functional reviews that are going on now, we will put together a business case and take that to Cat, who we will speak to as nicely as we possibly can. We will bow properly, and we will hope.

Q63 **Chair:** That business case will presumably include some progress, proving that you have hopefully saved money and prevented unnecessary spending.

**Professor Sir Ian Diamond:** Exactly so. Understanding the impact of the decisions and the evaluation in terms of improved policy and, therefore, improved funding seems absolutely critical to me.

Q64 **Chair:** Professor Diamond, this may sound an odd question, but do you consider yourself a Whitehall insider or an outsider coming in to help Whitehall improve?

**Professor Sir Ian Diamond:** I consider myself an outsider who is enjoying enormously the real dedication and professionalism of the inside of the civil service. If you were to ask me in a couple more years, I might



## HOUSE OF COMMONS

say, "I'm now an insider," but it's a bit like going to a village, isn't it? How long do you live in a village before someone says you're a local?

Q65 **Chair:** So, as an outsider to the village, then, do you feel that you are taken seriously and listened to?

**Professor Sir Ian Diamond:** Oh, I'm listened to, 100%. People have been very, very generous in asking questions and I can point to places that my advice—our advice—has been taken. And I praise enormously the dedication of my colleagues, who have had to be, if you like, a little bit more agile, a little bit more pacy and a little bit more radical in the way they do statistics, which I think has been to the benefit of the country.

**Chair:** So, learning both ways?

**Professor Sir Ian Diamond:** Yes, I hope so.

Q66 **Chair:** Long may that continue. Ms Hutchinson, you talked earlier about the interesting point of hard-to-evaluate areas, which we maybe sensed you could not perhaps say more about. But for the Public Accounts Committee, that creates a quiver of excitement, because some of us have been in Government or near Government, and it is really a big challenge when something is hard to evaluate.

Also, in questions the issue came out of the evaluation of the impact of reduced spending. Would you consider reducing spending a harder area to evaluate? And can you give us examples of where you have done that?

**Catherine Hutchinson:** Also, I would like to offer to share our top 10 areas with you, when they are able to be shared, for your own visibility of them.

The question, as I understood it, is this: how would you evaluate a reduction in a hard-to-evaluate area? Okay. For me, essentially an evaluation is a change. So, what difference has there been between two possible outcomes? What would have happened if you had that and what would have happened if you didn't?

For me, it is exactly the same method for introducing something, because you are comparing something with and without, as if you had it and then you took it away. So, all of the normal processes and methods—

Q67 **Chair:** So it is not actually as hard as all that?

**Catherine Hutchinson:** No. You could easily—one of the methods that we suggested during the spending review was to evaluate the scale-down. So, in the same way as you might look at—I don't know—whether a subsidy of 50%, 25% or 0% made a difference, you could scale something down and see if it had the impact you thought—

**Chair:** I partly asked the question because we rarely see that. So that is really helpful for us to continue to pursue as a Committee.

Thank you all very much indeed. It has been a really interesting conversation, on something that is very much meat and drink to us as a



## HOUSE OF COMMONS

Committee. And let's long hope that Sir Ian Diamond is listened to on this, because we do need more of this, as we know.

The transcript of this session will be up on the website uncorrected in the next couple of days and we will produce a Report, probably after the Easter recess or maybe during it, depending on how fast we can get moving. I thank you very much indeed.