



HOUSE OF LORDS

Science and Technology Committee

Corrected oral evidence: Innovation in the NHS: Personalised Medicine and AI

Tuesday 21 April 2026

11.50 am

[Watch the meeting](#)

Members present: Lord Mair (The Chair); Lord Booth; Lord Drayson; Lord Duncan of Springbank; Baroness Jones of Whitchurch; Baroness Nicholson of Winterbourne; Lord Patel; Lord Stern of Brentford; Lord Verjee; Lord Willis of Knaresborough; Baroness Willis of Summertown; Lord Winston.

Evidence Session No. 10

Heard in Public

Questions 110 - 118

Witnesses

I: Professor Raghiv Ali OBE, CEO, Chief Investigator and Chief Medical Officer, Our Future Health; Professor Sir Rory Collins FRS, FMedSci, FRCP, Chief Executive, UK Biobank.

USE OF THE TRANSCRIPT

1. This is an corrected transcript of evidence taken in public and webcast on www.parliamentlive.tv.

Examination of witnesses

Professor Raghib Ali OBE and Professor Sir Rory Collins FRS.

Q110 The Chair: Welcome to the second part of this morning's inquiry. We are very pleased to have with us Professor Raghib Ali, who is the CEO, chief investigator and chief medical officer of Our Future Health. We are also very pleased to have Professor Sir Rory Collins, who is the chief executive and principal investigator of UK Biobank. You are both very welcome.

The UK Biobank has been running for many years, and Our Future Health is, in a way, the successor generation that is enrolling more patients all the time. Can you give us some insight into the nature and purpose of the study, the types of data being gathered, what the programmes have achieved so far, and what they intend to achieve in the future?

We are interested to know how the two programmes complement each other and where they differ. How does Our Future Health look to build on UK Biobank's success? Perhaps, Professor Collins, you would start.

Professor Sir Rory Collins: Thank you very much. UK Biobank is a global research resource for discovery science in health. It is like the CERN Large Hadron Collider for physics or the James Webb Space Telescope for astronomy. It essentially allows the invisible to be made visible, for discoveries to be made that otherwise would not be made. It allows that to be done at very much lower cost, because the data are there for researchers to use, and very much faster than was otherwise the case.

UK Biobank was established by the Medical Research Council and the Wellcome Trust in 2003. I was asked to take it on as chief executive and PI in 2005. We were asked to recruit half a million men and women from across the UK, aged 40 to 69, largely healthy. With advice from experts across the UK and internationally, we developed a protocol for asking them all a series of questions about their environment, lifestyle and family history of disease; making physical measurements of them, including blood pressure, hearing and vision; and collecting biological samples that could be analysed at that time, for example, to do biochemistry measurements, such as cholesterol, haematology measurements around anaemia and white cells and red cells, as well as stored for analysis to be done in the future.

Those samples have now been genotyped and sequenced; they have undergone genetic sequencing, and metabolomics and assays have been done on them, so looking at the metabolites in the blood. Currently, a consortium of companies, along with matched funding from government, is doing measurements of thousands of proteins in those samples and making the data available. We have imaged 100,000 of the participants' brains, hearts and bodies and all those data have been made available to the wider research community.

We have this very large-scale cohort, with very deep information about their exposures to genetic, lifestyle and environmental risk factors and, because of the prescience of the MRC and the Wellcome Trust in setting this up more than 20 years ago, we now have long-term follow-up of an increasingly large number of health outcomes that allows researchers to determine why it is that one person does or does not develop a disease.

Perhaps the most important dimension of UK Biobank — to add to the scale, depth and duration that the MRC and the Wellcome Trust insisted on — was that we made these data available, with the consent of the participants, to academic and commercial researchers around the world, meaning that the altruism of the participants and the investment in the resource created the largest number of discoveries that have improved the diagnosis, prevention, prediction, and treatment of disease.

There are over 20,000 researchers around the world using the data. There have been nearly 20,000 peer-reviewed publications based on UK Biobank—5,000 in 2025 alone—and almost 1,000 patents based on UK Biobank. When we talk to people in industry such as Hal Barron, the previous chief scientist at GSK, he and others say that UK Biobank is transforming drug discovery. We are starting to see new targets being identified and new drugs being tested that we hope will soon come into clinical practice.

Finally, we can claim a major contribution to at least one Nobel Prize: Svante Pääbo got the Nobel Prize in Physiology or Medicine in 2022 for using the UK Biobank data to identify that Neanderthals were more sensitive to pain. You can see that now; it is still present in humans.

The accessibility of UK Biobank, which was the vision of the MRC and Wellcome Trust have, I think, made it a spectacular success, which other studies around the world are now following.

The Chair: Thank you. Professor Ali, how does Our Future Health build on UK Biobank's success?

Professor Raghieb Ali: First, I should say that a study such as Our Future Health is only possible because of the success of UK Biobank. We are very much building on the lessons learned and the innovation of UK Biobank.

The reason for Our Future Health being established goes back to the industrial strategy challenge fund and plan from 2017 to 2018, recognising that our current model of healthcare—which I very much recognise myself as a front-line doctor—is based on treatment of late stage disease, when people already have signs and symptoms of disease. Often, particularly in the emergency department, we see patients with heart attacks, strokes, cancers that have already spread, advanced dementia and so on, and it is too late for them to have a good outcome and it is bad for the system. It is very expensive. It is bad for patients, bad for population health and bad for the NHS as well.

Recognising that, one of the main reasons that Our Future Health was established was to see if we could establish the evidence base for a new approach to health in the UK. Instead of our having, in essence, a national sickness service, it is moving towards a real national health service, where we try to keep people in good health for as long as possible by establishing the evidence base for new approaches to risk or prediction of disease, early detection of disease, early intervention, and thereby prevention or at least delaying disease as much as possible. You expand healthy life expectancy and compress morbidity, which is obviously best for our population, for our patients and, ultimately, for the NHS.

The current system is really not sustainable, particularly with an ageing population, and it brings huge challenges in multi-morbidity. Since the pandemic, we have also seen increasing ill health in working age populations, particularly obesity and its complications and mental health disorders and their complications.

That was the background, and it was felt in order to tackle that and to establish the evidence base we needed a new prospective cohort study of 5 million people. The reason for that very large sample size was to enable us to look at diseases across all main parts of the population from age 18-plus; all ethnic groups, as the UK has become a very multi-ethnic society; all socioeconomic groups, including much larger numbers from more deprived populations than has been the case previously; and all parts of the UK. Again, most studies previously have been able to recruit only from certain parts of the UK. The size and diversity of the study would enable us to establish the evidence base for all parts of the UK population.

The other key new parts that are relevant to Our Future Health are that we agreed from the beginning that the study would be run in partnership with the NHS and that the evidence base that we were establishing would be something that could be implemented within the NHS. That would involve giving feedback to our participants about their risk of disease and running risk-stratified screening programmes, for example. That is why we have had that partnership from the beginning.

Thirdly, it was to be able to invite our participants to take part in recontact studies, including clinical trials, to establish the highest level of evidence possible, particularly randomised control trials. All our participants give consent to be recontacted for future studies.

Finally, with such a large cohort of 5 million people—about 10% of the adult population of the UK—it will give us a unique insight into the health of the whole population, particularly if we could get a representative sample from all parts of the UK.

It is a relatively new study. We launched recruitment in October 2022, so just over three and a half years ago. Fortunately, we have had a remarkable and unprecedented response from the UK public. Over 2.75

million people have joined to date. Of those, over 2 million people have completed the questionnaire that we asked them to do. Over 1.6 million people have attended an appointment to give their physical measurements and donate a blood sample. All have consented to have their linkage to their health records, NHS records, and to be recontacted.

We are continuing to recruit. We are very confident now, thanks to the support we have had from government, industry and charities, that we will reach that 5-million target over the next two and a half years. We have also just started our clinical research recruitment service, so we are starting to run clinical trials from this year. We have made a commitment in the Life Sciences Sector Plan to recruit up to 50,000 participants into clinical trials by 2030.

Q111 The Chair: It is a very impressive programme. We have more questions about that and about UK Biobank.

Professor Collins, this one is specifically for you. Could you tell us about the frontier of epigenetic sequencing and UK Biobank's work with Oxford Nanopore? How might that change our ability to develop personalised medicine from genetic data?

Professor Sir Rory Collins: This is a very exciting opportunity. More than 70% of all the large pharma and biotech companies around the world are using UK Biobank, but there are also 500 UK small and medium-sized enterprises, SMEs, that are using the UK Biobank data. UK Biobank data are helping to drive their mission: Genomics plc, a spin-out from Oxford University that Peter Donnelly set up, is using UK Biobank data to generate predictive scores, using polygenic scores that are already starting to be implemented into NHS practice and are helping to guide drug discovery.

We have been working with imaging study groups, including SMEs such as Perspectum, that can develop imaging analysis of the liver and can identify high levels of iron in the liver that may identify people at high risk who you can treat. With the proteomics in UK Biobank, we saw about a dozen companies come together to fund proteomics in 50,000 UK Biobank participants using a platform called Olink, which is a spinout from a Swedish university, measuring about 3,000 proteins.

They found that those data were valuable for drug discovery. Indeed, others found that it was incredibly predictive of disease. You could identify people who were going to develop dementia 10 years before using the proteins because the proteins are the common pathway of genetics, lifestyle and environment. They are even more predictive than genetics alone. As a consequence of that, those companies are now, along with the Government, funding doing the Olink proteomics on the whole cohort.

I think that leads into Oxford Nanopore. When I talked to Oxford Nanopore, I said the big problem is that all sequencing, and indeed epigenetic assays, are done using Illumina, an American company. To

help demonstrate the power of the Oxford Nanopore technology, which in many ways in principle is much better, it gives you much longer reads of the sequence, which helps you to identify mutations that are not picked up with the Illumina sequencing, so-called insertions and deletions, or repeats of the genome that can be associated with disease.

It also provides much better information about environmental changes to the genome, called epigenetics, so that you can see whether someone is or is not a smoker. That opens up the opportunity to look at lifelong exposures to different aspects that we cannot measure now, but we could measure them with the epigenetics.

My hope is that, with the Government funding for 50,000 samples being analysed with Oxford Nanopore, it will provide a real push to that technology being demonstrated to be of value and to being adopted. The more competition there is for different methods for sequencing and for epigenetics, the lower the cost and the more accessible it is.

Lord Patel: Excuse me for being obtuse. I think that what you said is quite exciting in science terms: lots of publications, even a Nobel Prize that showed that Neanderthals suffered from pain and experienced pain—that is great. What tangible benefits have been derived from patient care, considering that quite a bit of funding for UK Biobank—and I was involved at the time of establishing it through MRC—comes from taxpayers? Who holds the 500 or so patents that you described? What benefit is it to UK plc?

Professor Sir Rory Collins: You are quite right: there has been a substantial investment from the UK Government into UK Biobank. There has also been about £200 million from charity. Some £250 million has been leveraged from industry and now £40 million from philanthropy, making UK Biobank as valuable as possible.

In terms of tangible findings, through the imaging project in UK Biobank, 100,000 people have been imaged and now 60,000 of them are being repeat-imaged. Those imaging data at that scale—10 times bigger than ever done before—created an engineering problem. The engineers, with AI and machine learning, developed automated approaches to analysing those imaging data. The imaging protocols and those automated approaches to analysing the imaging are now already being incorporated into clinical practice. Memory clinics across the UK are using our imaging protocol and the automated approaches to analysis to speed up and improve the diagnosis of problems in people in memory clinics.

The cardiac imaging automated analysis that was developed using UK Biobank is now being used by about 1,000 hospitals across the world to reduce the time it takes to analyse the cardiac images, to free up radiologists and to have something that took 15 to 20 minutes done in seconds.

The Chair: Baroness Willis has a question. She is coming in online.

Q112 Baroness Willis of Summertown: Following on from that question, I want to ask about how you get the outcomes from the biobank and these other huge datasets—which are superb; I do not think anyone would question the value of them—into clinical trials and clinical practices, and then to the patient end of medicine.

Before I do that, I just ask about Oxford Nanopore and that link that you were describing. Oxford Nanopore now very much works with and within the Ellison Institute, where the IP will be owned by the Ellison Institute and it will be, in effect, a US funder. How does that link back to what we were saying about accessibility and how this will benefit the UK National Health Service?

Professor Sir Rory Collins: This is a question for the financial industry within the UK and the investor industry in the UK. There are other examples. Eliptica, which is a spin-out from the Crick, has developed a mass spec method for assaying samples in UK Biobank. It is a 50,000-person study that can demonstrate value and which I hope would become a larger UK company. We have seen AstraZeneca use the UK Biobank data to develop what it called MILTON. This used the first 50,000 proteomic assays to predict hundreds of different diseases many years before they occurred. We are starting to see that done in drug discovery. GSK is using the UK Biobank data to make major drug discovery targets and taking them into practice.

The other major observation from UK Biobank is the demonstration of the potential for polygenic risk scores to identify people who are at high risk of disease. Peter Donnelly, the founder of Genomics plc, which is a UK spin-out company that depends very largely on UK Biobank data, has shown that familial hypercholesterolemia, one of the most common single gene disorders that raises your risk of heart disease, is observed in about 0.3% of the population. With a polygenic risk score, according to Peter Donnelly, you can identify 10% of the population that have a risk equivalent to familial hypercholesterolemia. That is now being implemented within the NHS as part of the QRISK score that is used by GPs to identify who can benefit from lowering their cholesterol.

Rather than it being based on age and sex—men aged 50 get offered testing for statin therapy—we can use a polygenic risk score that will identify 10% of the population at a young age where we could offer them ways of reducing their risk of cardiovascular disease. That single test, that genotype test, that you can use to create a polygenic risk score can also be used to identify the few per cent of the women in the population—10 times more than have, say, BRCA1 or BRCA2 single gene disorders—who might benefit from early breast cancer screening.

For prostate cancer, the PSA test is a very poor test for identifying people at risk of prostate cancer. If you see a high value, it is difficult to know what to do, but if you combine that with a high polygenic risk score for prostate cancer, that might lead to different interventions compared with those with a low score.

Lord Patel: Can you describe the components of this polygenic risk score in prostate cancer apart from PSA?

Professor Sir Rory Collins: What the polygenic risk scores are doing, what we have seen—

Lord Patel: What are they?

Professor Sir Rory Collins: A polygenic risk score, unlike a single gene mutation—

Lord Patel: No, in prostate cancer, you said that, apart from using PSA as a test, which is not that reliable, you can use polygenic risk scores. So my question is: what are the other components of the polygenic risk score, apart from the PSA?

Professor Sir Rory Collins: The polygenic risk score combines variants across the genome of small effect.

Lord Patel: What are they?

Professor Sir Rory Collins: They are a lot of different variants across the genome. I cannot tell you which specific ones they are, but you combine a number of different variants across the genome, and the combination, based on UK Biobank data and now on other biobank data, can identify a few percent of the population that are at very high risk of developing prostate cancer.

A different combination of variants will identify a few percent that are at very high risk of breast cancer, cardiovascular and so on. Again, it has been shown that, with a genotype, you can develop polygenic risk scores for different diseases. Between about one-quarter and one-third of the population can be identified in the top few per cent—

Lord Patel: I have got it, thank you.

The Chair: Baroness Willis, had you finished your questions?

Q113 **Baroness Willis of Summertown:** I was going to ask the second part to this question. We are hearing some excellent examples, but can you say what the barriers are? We heard from one of our witnesses last week that one of the barriers was the link between the GPs and the data the GPs hold and how that feeds back into these larger datasets. I do not know whether one of you wants to comment on that.

Professor Raghieb Ali: To add to the first question, as Rory has described, there is a potential use of polygenic and integrated risk scores in a number of diseases. Where Our Future Health comes in is that we will be there to establish the evidence base for those risk scores in clinical practice and work with the NHS to see how they can be implemented. If they lead to better clinical outcomes and are cost-effective, that is the critical next step. It is a good example of how we are building on the

discovery science done by UK Biobank to then show how they can be implemented in our healthcare system.

We are very new, as I have said, having only run for just over three years. We have started our first clinical trial, which is a good example of how a resource such as Our Future Health can be used. It is an oesophageal cancer prevention study, working with Cancer Research UK and the University of Cambridge on a test that identifies people at risk of oesophageal cancer much earlier, something called Barrett's oesophagus. We did a first pilot study earlier this year and we were able to recruit thousands of participants into the study much faster than if they had been outside Our Future Health and with a much higher response rate. This is one example but, as I said, we have only just started.

On GP data, the key to being able to understand both risk factors for disease and outcomes of our participants going forward is to have data from multiple sources. That includes social determinants of health, lifestyle, environment and NHS records. One limitation that we have had to date, with both UK Biobank and Our Future Health, is not having access to the primary care records, where the majority of interactions take place. Many diseases are only dealt with in primary care. Mental health disorders, for example, are primarily dealt with in primary care, as is dementia and many other disorders. Without that data, we really have one hand tied behind our back in terms of the discoveries that we could make and developing the evidence base.

The good news, after many years of efforts from many people, including Rory, is that earlier this year a directive was signed by the Secretary of State that will allow consented cohorts. These are cohorts—UK Biobank, Our Future Health and Genomics England—where we have explicit consent from our participants to link to the GP records and where the Department of Health has taken on the responsibility as data controller for that data. We will be able to apply for that data later this year, with appropriate security safeguards in place and ensuring that our participants understand what they will be signing up to and what they will be linking to. If that application is successful, we expect—we hope—to have that data in our trusted research environment by the end of the year.

Baroness Willis of Summertown: Can I just finish up on that? It is just for my own understanding. That sounds fantastic, and you get a link through there, but do you have the infrastructure, particularly the computing backbone and infrastructure, to enable that to happen at speed? Another issue that we have heard about is that things are very siloed, and being able to break down those silos seems to be one of the biggest challenges.

Professor Raghb Ali: For Our Future Health, we have the infrastructure in place to receive that data and we are prepared to receive that data. We have been working with NHS England for the last year in anticipation of this announcement. The size of the database that we are applying for

is relatively small, for example, compared to genetic data, so the computer power required for that is relatively small.

Baroness Willis of Summertown: The GPs will have the training to be able to upload that data, will they?

Professor Raghb Ali: NHS England already has a copy of the data from the GP collection service.

The Chair: Professor Collins, are you optimistic now about the issue with GP data?

Professor Sir Rory Collins: The pandemic showed how valuable the GP data are. During the pandemic, a so-called COPI notice was used to allow primary care data to be made available to try to understand important determinants of bad outcomes with Covid. In UK Biobank, having access to the primary care data allowed researchers to identify particularly that ethnic minority groups and lower socioeconomic groups were at especially high risk from Covid, which was a very valuable use.

As Dr Ali points out, if you have access to the GP data, you substantially increase the number of cases of certain diseases—depression, dementia, asthma—that you can study. You more than double the number that you can study, and you can do so at an earlier age when they are more likely to be amenable to treatment.

The underlying problem is that, unlike the data on hospitalisation, cancers or prescriptions, which are controlled centrally within England, Scotland and Wales, the primary care data controllers are the 6,000 general practices around the country. Their focus is on trying to give care to their patients, not on data controllership. They do not want—this is what I am told—to have the legal liability associated with that. Therefore, the direction that has been issued, which is like a COPI notice by the Secretary of State, for access to the coded GP data for consented studies is a step forward, but it is a small step forward.

The question is: what would transform the situation? In my view, that would be by—as with the hospitalisation and the cancer data—centralising the data controllership of the GP data, taking away the liability from GPs and ensuring that those data are accessible using a consistent and high-quality approach, as occurs with NHS England at the moment and is already done in Scotland, where there is shared data controllership. There is an opportunity for a real transformation if we move from a directional, COPI-type approach, to centralising the controllership of the data.

To Baroness Willis's question about technology, the beauty of the GP data is that they are on only two mature IT systems. One was called Emis, now called Optum, which was sold to UnitedHealthcare, an American health insurance company, in 2023 for £1.2 billion, and one is the TPP system. They are two very mature systems; there are 6,000 data controllers, but two data custodians. If the centralisation of the data

controllership were with the NHS or potentially HDRS, you would open up the opportunities to use these data appropriately but effectively.

Lord Winston: Of course, one issue that arises in your programme is addressing a historical lack of diversity in large-scale cohort populations. Do you think that this will be a major problem with regard to the application of personalised medicine across different groups in the UK?

Professor Raghieb Ali: I am happy to start on that one. When we started recruitment for Our Future Health, we were committed to ensuring that we had better representation of underrepresented groups than had been possible previously. I mentioned at the beginning that 2.75 million people have consented. Of those, about one-third are from underrepresented groups, including ethnic minority groups, more deprived populations, parts of the UK with very poor outcomes that have not been represented in these types of studies before—coastal areas, more rural areas, central Wales, Scotland, south Wales and so on—and in terms of age. Most studies in the past have been from age 40-plus, and our study includes people from age 18 to 40. We now have nearly 1 million people in the study from those underrepresented groups.

We took a deliberate approach to increase diversity through additional outreach efforts with various communities that have been underrepresented. We have shown over the last three years that it is possible and that people from those backgrounds are interested in taking part in this research, as long as we explain clearly why we are doing the research, what the benefits are to them and to the health of the population, and the safeguards that we have put in place to protect their data.

We have made very good progress, but there is a long way to go. We still want to ensure that no one is excluded from taking part in our study and that all the different parts of the population can benefit from our research.

Lord Winston: What are the difficulties in getting a more diverse population? Are there specific problems there?

Professor Raghieb Ali: It depends on which groups you are talking about, but one of the main issues historically has been accessibility, in that they have not been invited to take part in these research studies in the first place. Many geographical parts of the UK have never had the opportunity or have never been invited to take part. It may also be because they are the wrong age—ie, younger adults. In certain ethnic groups, it is fair to say, as we also saw during the pandemic, that there are low levels of trust both in the healthcare system and in research in general; there are good historical and some contemporary reasons for that. It takes time to build trust with those communities. Those from more socioeconomically deprived backgrounds, again, may not have had the opportunity or been given the access.

We talk often about the inverse care law and I would say that something similar applies in research and in inverse research: those who are most able to benefit from research are also the least likely to take part. As I said, we have done a lot of outreach work across the whole of the UK, across all four nations. We have recruited hundreds of thousands of people from those groups. We are not complacent. We will continue to make those efforts over the next two and a half years in which we are recruiting participants. We will also ensure that we maintain engagement with those participants so that they understand the opportunities to take part in future research programmes as well.

Professor Sir Rory Collins: Certainly, in UK Biobank we went out of our way to try to increase socioeconomic diversity and other aspects of diversity, but diversity is broader than that. We want to understand the full range of exposures and their impact on disease, across all different diseases. When we established UK Biobank, we worked with researchers in other countries to establish similar cohorts that increased the range of exposures and diseases that could be studied. We have a study with the Mexican Ministry of Health of 150,000 people, which has allowed us to study the impact of obesity and diabetes at much higher rates than we see in the UK. That is, if you like, looking at where we are going to.

We set up a study of half a million people with the Chinese Academy of Medical Sciences in parallel with UK Biobank, where we are able to look at differences in risk factor exposures. A study that we did in China showed that they had much lower cholesterol levels than we see normally in the UK, which was associated with much lower rates of heart disease. It was that observation that led us and others to go on to test whether lowering the cholesterol level that was considered normal in the UK lowered risk, and it did. On that basis of that observation from a more diverse population, it has helped the UK population by ensuring that we are giving cholesterol-lowering treatments to people who would otherwise not have been given them because their cholesterol level was considered normal. It is a bit like saying that 10 cigarettes a day is normal—you need to study a much wider range of exposures.

We have also helped do that in the US, where I was the only non-US member of the panel that planned President Obama's All of Us study. That, again, is a very diverse, and differently diverse, population with a large Hispanic contribution.

- Q114 **Baroness Jones of Whitchurch:** I would like to look at the whole issue of data access and data security. The Government are establishing the new Health Data Research Service and both of you have expertise in large data segments. Presumably, you will welcome that service as a single point of access. What does it need to do to get things right, to make sure that data is accessible for researchers in the way that you would want but is also secure? Perhaps I could ask a supplementary to that, which is: what do we need to do to speed up access for researchers? Some researchers are waiting months or even years to have access to the data, particularly non-health sector researchers.

First, do you think it is the right model? Will it work? Will it help you? Secondly, how can we speed up access for researchers to make it more flexible and immediate?

Professor Raghb Ali: We certainly welcome the establishment of this new service. It is very early days, and I think it is fair to say that the new CEO, chair and board that are being established will need to be given some time to come back to us with what their priorities are for the coming months and years, but certainly having a single front door for UK health data assets will be very helpful for researchers.

From an Our Future Health point of view, currently we have to apply across the four nations to access data. Some of these applications can take, as you say, not just months but years to access the data, and that is the case for other researchers as well. First, reducing the time that it takes to access that data will be very important.

Secondly, standardisation of the way that data are classified would be very helpful across different datasets across the four nations. We talked about primary care data earlier, but there are many datasets across the UK in regional NHS secure data environments, not just the hospital episode statistic data, which we can access, but also electronic medical record data, for example, from secondary care and tertiary care trusts. In the primary care data, the extract that we can apply for is not the same as the full dataset that is stored within those that Rory mentioned earlier.

Again, if HDRS could bring all those datasets into a single access point and a single application process—not that they all need to be in a single data lake—it would speed things up hugely and avoid duplication. At the moment, we may have to apply to 20 regional SDEs in different nations with the same application, the same research questions and so on. That efficiency would be very helpful, as would interoperability between the different datasets.

Professor Sir Rory Collins: I think that HDRS is a fabulous opportunity. There is a real opportunity for a quick win if agreement can be made for central access not only to the cancer and hospitalisation data but, as I mentioned earlier, to the primary care data. I wonder whether HDRS could become the data controller for primary care data; it would not be government because, although it is a government company, it has its separate independent board. Therefore, it might be something that would be satisfactory to GPs, taking away their liability without them feeling that they are being told exactly what to do. That would open up access to primary care data.

Professor Ali is exactly right that we are all replicating things. We are all having to go to all these different siloed datasets. If we want to understand the determinants of a particular cancer, we need to have very deep information about the cancer. It is not just breast cancer, colorectal cancer or prostate cancer. We want to have the histopathology and sequence data so that researchers can find out what it is and what

exposures—genetic, lifestyle, environmental—cause some particular subtype of those cancers. These are all in different silos.

If HDRS, as Professor Ali says, could be a central point where you can have that access, it would reduce the replication of activity by researchers. It would speed up access and the adding of those data to these cohorts, which the participants have consented to make available for research, and I think that it would be something that would attract an enormous amount of interest and involvement from outside the UK—particularly pharma—to use those data to make discoveries that improve health.

- Q115 **Baroness Nicholson of Winterbourne:** Of course, once health data becomes linked to commercialisation, patients, quite understandably, have very deep concerns. How are you able to not counteract that but bring the patient onside? Is there any way in which you might do some test surveys and see what would give them confidence? Might it be better to build them in at the beginning so that the patient agrees what is right to be known? You might find that they are much more tolerant about privacy rights than outsiders believe. What are you doing on this and how can you reassure us?

Professor Raghb Ali: Our Future Health was set up, somewhat unusually, as a public-private partnership from the beginning. It has always been transparent to our participants and to various stakeholders that we have involvement from the life sciences industry, which has made a significant financial contribution to the programme of around £170 million. However, they still have to apply to access the data in the same way as any other researcher, it still has to be clear that there is public benefit from the research that we propose to do and they have to go through the same access process as academics, the NHS or other researchers.

What we have done with our participants and the millions of people who have joined is to ensure that they know this from the beginning—it is transparent about the fact that we work with industry—and that they understand why it is important that we work with industry. There was a very good example, of course, during the pandemic, where the collaboration between Oxford University and AstraZeneca led to the development of a vaccine in record time with distribution across the country. That applies to most medicines; there is collaboration between academia, industry and the NHS that allows new medicines and new diagnostics to come to our patients.

We have done surveys on our participants and on the general public about their attitudes to industry both funding Our Future Health and accessing the data from Our Future Health, and the vast majority of our participants are, as you would expect, very satisfied that funding has come in and that data will be used by industry. Over 90% of our participants say they are either very or fully certain. I am happy to share that data if it would be of interest.

We have also asked the general public, and the numbers were higher in the survey that we did. Even in the general public, the majority of people understood that it was necessary to work with the life sciences industry so that we could get these new diagnostics and new treatments for patients.

Baroness Nicholson of Winterbourne: There is a difference, is there not, between public benefit and public service? The NHS is a public servant in that sense. The deduction as to what is a public benefit may not be the same as the assumption that the patient believes.

Professor Raghb Ali: I am not sure. Maybe you could repeat the question or clarify.

Baroness Nicholson of Winterbourne: The patient themselves, or the group of patients, may not agree with the public benefit that you have defined. How do you deal with that?

Professor Raghb Ali: In terms of the people who have already joined Our Future Health, I have already explained about the general, high-level benefit to them, potentially, from new treatments and diagnostics.

For individual studies—for example, if there was a new interventional study we were doing within Our Future Health—participants would have to consent to that again as a new consent. They would explicitly say that they agree to take part in a study for a potential new treatment or diagnostic test. I hope that partly answers your question.

The Chair: I am afraid we have to move on.

Q116 **Lord Drayson:** Can we turn to the aspect of how UK Biobank and Our Future Health contribute to UK economic growth? Professor Collins, you very helpfully compared UK Biobank to CERN; there is no doubt that UK Biobank has been a wonderful contributor to the wider public good, but CERN is funded by 23 countries and UK Biobank is funded by one. You mentioned that 1,000 patents have arisen. You will be aware that the most valuable patents are life science patents, so a conservative estimate of the value of those patents is in the billions of pounds. Who owns those patents?

Professor Sir Rory Collins: UK Biobank does not hold on to patents from work that comes out of the research on UK Biobank. It was decided at the beginning that, in order to encourage use of the resource and to turn the data into discovery, we would not do so. I just mention that it is not solely funded by the UK Government. There has been £250 million of industry investment.

Lord Drayson: I am sorry to interrupt you. We are short of time. I think that I am right in saying that the majority of the funding comes from the UK taxpayer.

Professor Sir Rory Collins: Absolutely. The UK Government have made a substantial investment and indeed have put another £125 million into—

Lord Drayson: To put you both on the spot with the time we have available, you will be aware of the funding pressure that the research councils are under. In fact, the Science and Technology Facilities Council, which contributes to CERN, is seeing a cut and redundancies as a result of these pressures. Do you think that it is sustainable for the UK to continue on the basis of providing economic value for free to other countries, or do you think that more should be done to ensure that the value that is undoubtedly generated by the fantastic innovations that you are generating flows back to the UK taxpayer?

Professor Sir Rory Collins: I think that the investment that the UK Government have made in the new facility in Manchester, which will allow us to provide samples much more rapidly for assay and analysis, will engender even more investment from outside the UK in turning those samples into data, so that we will have a much bigger proportion of the resource funded in that way. We can support SMEs in the UK, as I described earlier, and we can work particularly with companies like GSK and AZ, which are UK companies that are using the data enormously. As I see it, the role of UK Biobank is to make the cake bigger. The slice of the cake that the UK receives also gets bigger, even if others benefit as well.

Professor Raghieb Ali: For Our Future Health, the condition for the initial grant that came from Innovate UK of £79 million was to secure an additional £160 million from industry, a 2:1 match, which we were able to do. That was the initial investment into the programme.

Going forward, in terms of the research that will be conducted within the Our Future Health resource and particularly clinical trials, as I mentioned, we anticipate that even if 1% of our cohort—which will be 50,000 people—take part in clinical trials over the coming years, it will generate billions of pounds in GVA into the UK economy. This has been estimated by the ABPI. Clinical trials are of course good for our patients because they bring new treatments and new evidence for them, but they are also good for the economy. They generate many jobs and a lot of inward investment into the UK. This is one of the ways that we will be able to make a significant impact.

Lord Drayson: It would be helpful if you could write to us and provide the economic analysis that backs up that claim.

Professor Raghieb Ali: Of course; we are happy to. It has been reviewed by Treasury already.

The Chair: Thank you very much.

Q117 **Lord Duncan of Springbank:** I would like to understand what AI can offer. You have an extraordinary resource available, but can you talk us through exactly what AI can do and what guardrails we need to be alert to in the use of AI in that area?

Professor Sir Rory Collins: I am not an AI expert but I think that there are two issues related to AI. One is technical. Increasingly, resources like UK Biobank and Our Future Health—all of us—are sitting on cloud-based platforms that democratise access for researchers around the world because you do not need to have a big computer; you can come and work on the computer on the cloud-based platform. The underlying issue is that those platforms do not typically have the compute power or indeed the tools that would allow you to do the hardcore AI that you would like to do.

We have been working with the UK high-performance computing systems Dawn and Isambard, and the Novo Nordisk Foundation-funded Gefion in Denmark, to look at having UK Biobank data in a secure environment on those high-performance computing systems in order that the tools and the compute are there to do AI.

The second issue is around data security. If you develop models using AI on data, there is a risk of the data being what is called “memorised” within the model. If you then take the model weights off your secure environment, the data have gone. The question is: how do you allow AI to be applied to data, to build models that can help to improve health, in a secure way? In the All of Us study in the US, in our study and, I think, in Our Future Health, the approach is that the model weights cannot leave the platform. That avoids this risk of memorisation and data being taken off.

There are technical and data security issues that need to be addressed. However, AI requires data and that is what these cohorts are providing in spades. They are providing the deep, large-scale, long-term data that AI needs. To Lord Drayson’s question, that attracts the AI community to come to the UK because we have the best data. Only yesterday, Boehringer Ingelheim announced that its new AI enterprise will be located here in the UK, because this is where some of the top AI scientists are but, more importantly in my view, where the best data are for applying those technologies to understand how to prevent and treat disease better.

Lord Duncan of Springbank: If you are only using your platforms, does that then create a bottleneck for utilisation? There will clearly be a problem in that scenario.

Professor Sir Rory Collins: Yes, the platforms are a bottleneck, which is why, in collaboration with the Turing, Dawn, Isambard and the equivalent in Denmark, we are looking at having a set of the data in a secure environment under our control where you can have the compute power and the tools to do the AI analysis, but you also have a secure process that does not allow the memorised weights to go off. Bear in mind my lack of expertise in AI, but this is the guidance that we are being given by experts in this area.

The Chair: Lord Patel, I think that the question has been answered.

Lord Patel: It has been answered.

The Chair: Thank you. Our final question comes from Lord Willis, who is online.

Q118 **Lord Willis of Knaresborough:** It has been an incredibly interesting conversation this morning. Our inquiry will make conclusions and recommendations to the Government around innovation in the NHS, personalised medicine and AI, and you both lead large and very successful population health research programmes that can provide the insights that we are very keen to see benefit the NHS. When we produce our report, what recommendations would you make to the Government to get the most out of these extraordinary resources? We are looking for the three major things that you would say to the Government in this report, "If you do this, the Government and the country will develop and will benefit enormously".

The Chair: Professor Ali, what are your top three?

Professor Raghib Ali: I know that we are short of time. The first, which we have covered already, is about data access. Data is the fuel, as we have heard, for AI. We have the best data in the world in many ways for health research. Our Future Health and UK Biobank are two globally unique resources, but we need to be able to access that data more easily than is currently possible and we hope that HDRS will help with that.

The second thing is the regulatory environment. We want to work with the MHRA and other regulatory bodies in the UK to ensure that regulations are proportionate to what we want to achieve through a resource like Our Future Health. If you want to move towards a model of precision public health prediction, the regulations currently are not set up for that. They are mainly based on diagnosis. We want to work with the MHRA to ensure that we have a safe but appropriately permissive regulatory environment to maximise benefits for the health of the nation.

The third, and it is particularly relevant to Our Future Health, is the ability to ensure that the discoveries that we make in Our Future Health can be implemented within the NHS and that we have the workforce capability and the models for service delivery in place. Our Future Health in many ways could be seen as a pilot study—it is a very large pilot study, 5 million people—for the whole of the UK population, and what we discover within Our Future Health we want to be implemented across the population for the benefit of the whole country.

Professor Sir Rory Collins: Thank you very much for that question, Lord Willis. UK Biobank at the moment is studying the determinants of disease; I would like it to study the determinants of health and healthy ageing. We only study disease because we link to health records, which are about disease. We could more than double the value of UK Biobank by bringing all the participants back. We would then be able to look at how their life exposures have changed and how that relates to disease in the longer term but, much more importantly, we could assess their health

and their healthy ageing. Why is it that people age at different rates in terms of cognition, hearing, vision, bones and joints? If we had quantitative measures of health for hundreds of thousands of people, rather than 20,000 with some particular disease, we would massively increase the ability to study the determinants of healthy ageing. That would be my number one.

Secondly, like Professor Ali, I think that sorting out the GP data is critical. At the moment, as I mentioned, you have 6,000 data controllers who, I think, do not want to be data controllers. They want to focus on looking after their patients. They do not want the legal liability and the risks associated with that. Maybe there is some relatively minor legislation that could move the data controllership centrally so that GPs are reassured that it will not be used against them, but we can all be reassured that the data will be used in an appropriate way to help improve not just research but health delivery and the delivery of the 10-year plan.

With that second one, if HDRS was made the data controller—perhaps along with the secondary care data as well—I think that we would ensure that HDRS was a huge success and it could free up the other siloed data, enable much more research to be done and, again, help to deliver the 10-year plan, because the NHS could be a client of HDRS to help it understand how to deliver care better and consistently across the country.

The Chair: And your third?

Professor Sir Rory Collins: That was my third.

The Chair: Thank you both very much indeed for answering our many questions. We appreciate it very much. It has been extremely informative. We are very grateful to you both for coming and thank you for waiting so patiently before we started this second session. That brings this session to a close.