

HOUSE OF LORDS  
MINUTES OF EVIDENCE  
TAKEN BEFORE  
THE SELECT COMMITTEE ON SCIENCE AND TECHNOLOGY  
(SUB COMMITTEE II)  
**GENOMIC MEDICINE**

WEDNESDAY 16 OCTOBER 2008

PROFESSOR RORY COLLINS, PROFESSOR ANDREW MORRIS,  
PROFESSOR DAVID PORTEOUS and PROFESSOR JULIAN SAMPSON

Evidence heard in Public

Questions 475 - 528

USE OF THE TRANSCRIPT

1. This is an uncorrected and unpublished transcript of evidence taken in public and reported to the House.
2. The transcript is not yet an approved formal record of these proceedings. Any public use of, or reference to, the contents should make clear that neither Members nor witnesses have had the opportunity to correct the record. If in doubt as to the propriety of using the transcript, please contact the Clerk to the Committee.
3. *Members* who receive this for the purpose of correcting questions addressed by them to witnesses are asked to send corrections to the Clerk to the Committee.
4. *Prospective witnesses* may receive this in preparation for any written or oral evidence they may in due course give to the Committee.
5. Transcribed by the Official Shorthand Writers to the Houses of Parliament:  
W B Gurney & Sons LLP, Hope House, 45 Great Peter Street, London, SW1P 3LT  
Telephone & Fax Number: 020 7233 1935

WEDNESDAY 15 OCTOBER 2008

---

Present

Broers, L  
Colwyn, L  
O'Neill of Bengarve, B  
Patel, L (Chairman)  
Perry of Southwark, B  
Taverne, L  
Warner, L

---

Witnesses: **Professor Rory Collins**, Chief Executive Officer and Principal Investigator, UK Biobank, **Professor Andrew Morris**, Professor of Diabetic Medicine, University of Dundee and Chairman of the Generation Scotland Scientific Committee, **Professor David Porteous**, Chair of Human Molecular Genetics & Medicine, University of Edinburgh and Founder of Generation Scotland and **Professor Julian Sampson**, Head of Medical Genetics, University of Cardiff, gave evidence.

**Q475 Chairman:** Good morning, gentlemen and good morning to the members of the public and for those members of the public there is information about the interests of the Members of the inquiry in the papers on the desk there. Can I welcome our witnesses today; we are very grateful to you all for taking the time and trouble to come and help us with our inquiry. The procedure will be that if any of you want to make an initial brief comment or statement, please do. Have you agreed that one of you will handle the question, divided up, or are you going to be individuals? I had assumed you would all be individuals.

**Professor Porteous:** Yes.

**Q476 Chairman:** Take the questions as they come and anybody can chip in. You will be aware of some questions and there might be supplementary questions as the Members feel them appropriate to ask. Do any of you want to make any initial statements or comments?

No. We will then go straight into the questions and let me start. First of all, what was the trigger that created the UK Biobank and Generation Scotland and what do you think is going to be the value in the medium-term and long-term of these biobanks in terms of public knowledge and healthcare?

*Professor Collins:* I can speak for UK Biobank.

**Q477 Chairman:** When you speak for the first time would you say your name for the record?

*Professor Collins:* My name is Rory Collins; I am the Principal Investigator and Chief Executive of UK Biobank, so I can speak for UK Biobank. UK Biobank is, I think, quite different from Generation Scotland; it is more of what I would call a classical population study where we are looking at all of the different kinds of risk factors for disease – the way people live, the environments in which they live as well as their genetics and seeing how those different factors can interplay to determine whether or not one person does get a disease or does not. It is much like, for example, Sir Richard Doll's study of smoking in 50,000 British doctors, which showed over a 50-year period the full effects on health of smoking – not just on lung cancer but many other types of cancers, heart disease, lung disease, etcetera. So we are not just interested in genetics, we are looking at the whole range of exposures. There was this realisation over the last ten or 15 years that many of the risk factors that we were looking at need much larger studies to understand their full effects and, in particular, how they interacted with each other. So Richard Doll was very lucky; he was looking at smoking, which produces very big effects. But for genetic factors and for many other factors their effects are quite modest, particularly when you look at how they interact with each other you need very much larger studies than were customary in the past. We have demonstrated that by combining smaller studies, for example smaller studies in the association of blood pressure and disease, which showed that blood pressure was much more important as a cause

of heart attack and stroke – not just in middle age but in old age – and indeed the effects of blood pressure continue down throughout the normal range and that has had a very big impact on, for example, the Food Standard Agency’s policy on reducing salt in processed food in order to lower population blood pressure because we now know from putting together these studies that lowering blood pressure in what we call the normal range will lower the risk of heart attack and stroke. So there is that realisation that we need much larger studies to understand the full effects of these different factors that led to studies like UK Biobank.

**Q478 Chairman:** So is the size of the cohort adequate?

*Professor Collins:* No. The bigger the study the more valuable it will be in terms of determining the ill effects of different risk factors. The combination of different studies that I have mentioned, the prospective studies, involved a million individuals and even there when you started to look within different circumstances at different ages or when you looked at the combination of different risk factors you start to run out of information because it is not the number of people in the study that actually gives you its power to study the risk factors, it is the number of people unfortunately who develop any particular condition, and even for the commoner conditions half a million people will be barely adequate. For the rarer conditions we will have to wait a long time until enough people have developed the condition of interest, whether it be joint disease, dementia, whatever, to be able to study it really reliably. But it is a very good start and there are a number of such studies going on around the world so that by combining the data from those studies we will be able, in an anonymised way, to unpick the risk factors to a greater extent than we can in any one study.

**Q479 Chairman:** So are these studies in other countries with which you might amalgamate the information all collecting a similar type of information?

**Professor Collins:** Similar but not identical. There is a programme going on called P3G, which is trying to harmonise different studies of the association of risk factors in genetics with disease.

**Q480 Chairman:** In case we do not all understand the abbreviation.

**Professor Collins:** P3G is about looking at studies of genetics and other risk factors in combination. For example, there is a study of half a million people in China that has recorded similar information about people's risk factors and exposures, which will be very interesting because in a way the Chinese are where we came from in terms of our cholesterol levels – they are much lower; their body mass index, their weight is much lower. So we can look at the lower extremes of the exposures that we cannot really state terribly well in Britain.

**Q481 Chairman:** What is the size of the study in Japan?

**Professor Collins:** In Japan? I am not sure of what studies in Japan you are referring to.

**Q482 Chairman:** Perhaps somebody else might help?

**Professor Porteous:** David Porteous, University of Edinburgh and Generation Scotland. To the best of my knowledge there are many countries that are starting to think about setting up similar sorts of studies but there is not one which compares with the state that the UK Biobank or the Kaduri Study in China has reached, in Japan.

**Q483 Chairman:** So what is the difference between Generation Scotland and UK Biobank?

**Professor Porteous:** I think there are important differences. There are also important similarities and, more importantly, it is the synergies that we have been looking at. You might say you have just heard that 500,000 participants of the UK Biobank is insufficient, or we would like more is perhaps the way that we would put it, and you might ask what can 50,000 Scots tell us if 500,000 UK citizens cannot? The crucial difference is that rather than

asking for single members of the public to come forward to the UK Biobank we are asking families to come forward to work in Generation Scotland. We are doing that because we are primarily interested in the way that genetic risk factors are involved in determining not just disease end points but risk factors that are associated with disease trajectories. So we have talked a little bit about cholesterol levels and that is something that you can measure and measure over a range but the question is, what influence do genes have on setting that level and modifying it on their own and in relation to the environmental exposures? We would like to know whether or not high cholesterol in a mother or father is transmitted and, if so, to what extent to your offspring; and we would like to be able to see the differences in the exposures of the parents in their lifetime, how they compare to the exposures of children, as they grow, through life. So that is why Generation Scotland has those sides to it: the generations, our study is to look at the effects as it passes from one generation to another; and it has a trans-generational component to it, anticipating the fact that research we do today is going to be of the most value to the next generation. That is the kind of timeframe that we should be thinking about.

**Q484 Chairman:** What is the size of the cohort?

**Professor Collins:** We are aiming to collect 50,000 individuals in one of the three studies that we are doing under Generation Scotland. I will briefly mention the other two, just for clarity. We have, as I say, three studies. One of the studies is simply to collect a very large number of well controlled subjects and that we have done through the Scottish National Blood Transfusion Service. So we have collected 5000 samples from fully consented blood donors and have from them blood which we can use to do the genetic tests, but also plasma from which we can do bio marker tests. So that will be, if you like, a general population-wide control sample set that anyone who has appropriate ethical approval can come and use as a reference centre. We also have a smaller set of individuals – 2000 we are collecting – where

we have not just parental origin but also grandparental origin and here we want to look more at the profile of the Scottish history. So we are selecting those individuals from a number of different discrete geographical locations where we have grandparental history – born in the same region – and we will be able to ask are there differences even within Scotland in terms of the profile of genetic factors and their relationship to health trajectories. That material will also include collections of samples that will allow us to establish cell lines. The importance of that is that although your emphasis is on genomic medicine and I think you are primarily thinking about the impact of DNA-based genetic tests, really we are talking about a whole suite of technologies that allow one to draw lines between the genes, the products of the genes, the R & E, the protein, the metabolites and the cell building blocks of the cell, all the way on to the functioning of tissues and organs and the whole body. That we see as the clinical end of the spectrum. So to fill in those gaps we need to be able to have material that will allow us to ask questions not just about genes but the products of the genes, so this study helps us to do that. But the main study is the so-called Scottish Family Health Study – and I am sure that Andrew will add to this – and the essence there is that we want to study families, we want to study the segregation of genetic markers in those families and see how they correlate with what we call quantitative traits – that is, measurable traits that we can relate to disease processes. By doing so in families we are able to add statistical power that will compensate for the relatively small sample size and in general terms if we get the family collection right we will have more or less equivalent statistical power to a study of ten times that number. The reason we are doing the family-based study in Scotland is because we can do it and because of various other structures within Scotland – the National Health Service for Scotland, the record linkage system, the fact that we have a relatively stable population so that we can follow over the generations makes a family-based study possible. To do so in the breadth of England and Wales is doable but more complex and challenging in terms of being

able to do the family history work and to keep track of those individuals because it is a more mobile population. I hope that gives you a feeling for the contrast.

**Q485 Chairman:** Thank you, it does. Professor Morris, did you want to add anything about the study?

**Professor Morris:** Andrew Morris, University of Dundee and Generation Scotland. I think we are in a rapid transition phase where we are going through a discovery of rare variants with a large effect in terms of individual clinical impact to the discovery of common variants of very modest effect, which are much harder to find and may have much less individual impact, yet can yield great insights into scientific discovery. To do this I think it is clear that we need very, very large studies to be able to have the power and the certainty to tease out the modest clinical impact that many of these genetic variants have. I think that is the basis of these large studies in Generation Scotland and UK Biobank. You may know that it is the sixtieth anniversary of the NHS this year; it is also the sixtieth anniversary of study called the Framingham Study in the United States, which in 1948 recruited only 5000 family members, yet the results of Framingham from those 5000 family members have yielded immense insights into classical phenotypic receptors, such as blood pressure, cholesterol, smoking, high blood sugar, and we still use data derived from those original 5000 patients in clinical practice today. So, for example, if you go to your general practitioner he may calculate your future risk of cardiovascular disease and that risk calculator is based upon those 5000 US residents. We use it globally. So I think the origins of these large studies are to be able to adopt the same approach as those original Framingham investigators but in the genomic era where we are looking at very small effects in large populations, hence the numbers are so important.

**Q486 Chairman:** So the cohort of these studies is likely to yield an understanding of linkage between genetic and common diseases. You are nodding, Professor Porteous.

*Professor Porteous:* I agree.

**Q487 Chairman:** Professor Sampson, did you want to add anything?

*Professor Sampson:* Professor Sampson from the Cardiff University. Another form of biobank of course is to bank samples from individuals who are already affected by specific diseases and this is complementary to the long-term prospective studies that we have already discussed. One such initiative in Wales is the Wales Cancer Bank, which is collecting both tumour tissue and blood for blood DNA extraction from instant cancer cases across Wales. This is a project funded primarily through the Welsh Assembly Government and charities in Wales. This is an ongoing project to create a resource for genetic and allied research into cancer. One difference from the initiatives we have heard about is that the cancer tissues themselves constitute an important element of the biobank and will enable the schematic changes, the changes in the genes that occur that drive tumour genesis to be examined as well as the genes that potentially predispose individuals to cancer through the analysis of the constitutional samples. Of course there are many other similar cohort studies underway in relation to other common diseases that may not be represented in studies such as Biobank or Generation Scotland because of recruitment issues, for instance in relation to psychiatric disorder because those patients are unlikely to take part in these large population-based studies.

**Q488 Chairman:** Thank you very much. Could you remind us, Professor Collins, what is the current budget of Biobank UK?

*Professor Collins:* UK Biobank is recruiting 500,000 people aged 40 to 69 and the budget for the recruitment phase of Biobank and establishing the systems for looking after the resource,

particularly of biological samples but also the data is £62 million, coming primarily from the Medical Research Council and the Wellcome Trust, but with funding also from the Department of Health and from the Scottish Government.

**Q489 Chairman:** So that is the initial funding; how much more funding will be required when you get to the stage of trying to find an association of common diseases?

**Professor Collins:** It depends how good you want the study to be. The more detailed the phenotyping the more detailed information you know about the participants in the study the more informative it will be about the range of diseases. So we are looking at ways in which we can get even more detailed information about at least a substantial subset of the participants – perhaps 100,000 or so of the participants in the study – which will allow us to look in even greater detail about the relationship between a range of exposures.

**Q490 Chairman:** So how much extra funding would be required?

**Professor Collins:** The extra funding required to do what we are doing at the moment, which is essentially once recruitment is finished, to continue the follow-up of the participants and to make the resource available for people to use, is probably of the order of a couple of million per year. The way in which the resource has been set up, we have put a lot of money into an automated sample storage facility which means that when people need to use the resource they will be able to do so very quickly and at low cost. The money has been spent upfront, if you like, to build an easily accessible resource. So for researchers who want to use the resource the cost of doing so will be very low, but of course the actual experiments that they do will depend on the particular type of experiment they want to do and that would obviously be subject to scientific peer review. One comment I would just like to make is that the way in which this study is designed, the primary way in which it will be used is to wait until people develop a particular condition. So in ten years' time, perhaps 10,000 people have had a heart

attack and you would then take the data and the samples from those 10,000 and from 10,000 similar people within the resource and you would then analyse those samples and those data in exquisite detail. The reason I am stressing that is that we do not measure everything in all half million; we measure the things that really matter in the people who are most important.

**Q491 Lord Broers:** How will findings that you get from biobanks be translated into clinical practice? Are there structured pathways facilitating translation?

*Professor Porteous:* I think that this is something which has been, certainly so far as Generation Scotland is concerned, recognised by a number of different groups that are involved in trying to do that translational research, solve the translational research problem. What is the translational research problem? I think there is a recognised gap between the basic science and getting initial understandings about relationships between genes and genomes and health and putting that into practice. We have heard a little bit about how we have done that in the past in terms of the high penetrate single gene disorders but this is a very different matter altogether. So this needs to be a partnership between the researchers, the biobanks, the National Health Service and industry, and industry clearly plays a very important role in this because the one thing that they can do is to take assays and develop targets and from those develop molecules that become drugs and test and have those for clinical use. But to do so effectively is becoming increasingly challenging and so many, many pharma are saying, “Actually the real problem lies in the number of molecules that fail in the clinic.” The reasons for those are numerous but one of the major problems is that some of the molecules that look as if they work well in the laboratory or in animal studies just do not transform into things that work well in the clinic. So we have, for example, in Scotland recently had a very large level of investment from WYE Pharmaceuticals who have come in and said, “We recognise that Scotland has the capacity to help bridge that gap,” so we have set up our specific translational medicine research consortium, which is half way placed

between pharma and academia, in order to try to better understand why drugs might fail in the first place and to succeed more effectively in taking new drugs through to market. Andrew might want to add a little bit to that in terms of how these relationships might work.

**Professor Morris:** I think the definition of translation is important. Cooksey defined two levels of translation. Type one is from bench to bedside and type two is actually getting known evidence into clinical practice from paper into practice. Sydney Brenner suggested we should have three types of translation in that observations in the population can actually lead to bi-directional translation and lead back to scientific discovery, and I think the biobanks are uniquely placed to contribute to each of those three types of translation.

**Q492 Lord Taverne:** I did not quite gather the difference between the first two types. One was to bedside and the second was into clinical practice?

**Professor Morris:** We know that there is a lot of existing evidence which is not actually implemented in the clinic, so it is to see something through right from discovery right through to clinical practice. Type one is from the bench to evidence and type two is from evidence to actual implementation. So it is a two-stage process.

**Q493 Lord Broers:** How is this translational research funded? Is it a fraction of the overall biobank funding and, if so, what fraction is that or is it separately funded?

**Professor Morris:** I think the phrase translational research has become a major priority in terms of UK healthcare funding priorities and with the creation of OSCHR – I believe you had Sir John Bell, who would have told you about OSCHR – we have actually pooled resource to try and drive translation and concertina the timeframe from initial scientific discovery to clinical implementation. I think it is currently a focus – whether it is sufficient funding I would not like to say.

**Q494 Lord Taverne:** When UK Biobank was started there was some criticism about the large scale of the funding required and also it was thought the whole procedure was rather slow – slow in recruitment and being established. To what extent were these criticisms valid and how far, in so far as they were valid, have they been overcome?

**Professor Collins:** As Chief Executive of UK Biobank I think that is directed to me. I would not like to comment on whether or not criticisms were valid; I do not think I am qualified to comment. In terms of the progress of UK Biobank I think with a budget of £62 million it is appropriate to spend time thinking how you are going to spend it. In particular, to consult very widely, both nationally and internationally, about what questions to ask people about their lifestyle, what measurements to make and what samples to collect and how to collect them in order that analyses can be done that have not yet been invented. As I mentioned earlier, the idea is that one will go back to these samples in ten, 15 and, if I am Richard Doll, 50 years' time and analyse samples using methods of analysis, measuring things that have not been invented. So a lot of work went into thinking about that; also about how to recruit individuals efficiently into this study in an appropriate way – how to find them, how to invite them, how to see them go through the visit and, as I mentioned earlier, to store the samples in a way that meant that it would be possible to find a particular individual's sample in ten, 15 years' time. So it took some years before we were ready to start. We then piloted very carefully what we wanted to do and modified that in the light of our pilot experience and we started the main recruitment in April 2007; we recruited the first 100,000 people by April 2008. We are at about 199,000 as I sit here now, so we will be 200,000 in October 2008. Our schedule for recruitment was to complete half a million people by the end of 2010 and we are about six months ahead of schedule, on budget. I think that is an appropriate way of using such substantial funds. The National Institutes of Health also looked at setting up a similar study in the US. They estimated that it would cost one thousand million dollars and have not

been able to do it, partly because of the huge cost but partly because they do not have the luxury of doing a study within a place like the UK where we have a National Health Service, which can really facilitate such studies.

**Q495 Lord Taverne:** What about longer term funding? You mentioned that some of the follow-ups would not take a lot of funding, but experiments would be much more expensive; and you also mentioned that the bigger the sample the better – it depends on what detail is required. What do you feel about the long-term funding; do you think it is secure? Are you worried about it? What are your main concerns about long-term funding?

**Professor Collins:** As a researcher one is never secure about funding until the grant has been given, but the fact that there has been such a serious investment both scientifically and financially in developing this resource because of the potential value of a resource of this size. It is, in terms of its size and detail, both in terms of questions, measurements and the samples, unique; there is no other study on this scale with this detail in the world. That is not to say that other studies are not complementary to it and have different aims, but it is a unique study. So I think that should help to get long-term funding. The amount of long-term funding required to actually maintain the resource, that is particularly to follow-up the individuals, is comparatively small. Even the cost of using the resource in the way I have described, where you only analyse samples and data from the small subset of individuals that are of particular interest for that particular condition means that the cost of doing an experiment are relatively low, and of course you are doing it in five, ten, 15 years' time when the cost of assays that we want to do now are high but as technology improves you can do more; it costs less to do; you are doing it on a focus sample so the costs of doing the experiments are relatively small.

**Q496 Chairman:** Are we at the same time investing in these technologies that you will require?

**Professor Collins:** The technologies in terms of analysing the samples, there is a huge investment going on. Five years ago the idea that you would do what is called a whole genome screen on thousands and thousands of people would have been considered fantasy but now it is routine. The Wellcome Trust is I think doing 300,000 genetic markers – or a million genetic markers actually on tens of thousands of individuals. Five years ago, as I say, that was fantasy. In five years' time we will not even be doing that, we will probably be doing sequencing on tens of thousands of individuals. The costs are falling rapidly in this technology; the ability to get information is increasing greatly. What we have not had until now, until these kinds of studies were being built, is actually the resources on which to apply this technology. That is what we are building.

**Q497 Chairman:** I am tempted to ask, with all this knowledge that you will gather, can you describe what the scene would look like to, let us say, some young person going to a clinic and asking, “What are my risks?” What would it look like?

**Professor Porteous:** I will take the risk of starting to try and answer that. What I would say is it is not like I imagine it will be; it is going to be different from that, but I will try and give you a vision of what might be happening. There is no question that the technologies for analysing gene sequences and the products of gene sequences very importantly, is tumbling and there is a huge technology push to improve those further – there is no question about that. If you simply take, as we have heard, the initial cost of setting up the sequencing of the human genome – 3000 billion dollars for 3000 million bases of sequence – now we are just at the point where somebody is about to win the price for sequencing a human genome for \$1000, from start to finish. So the costs are coming down dramatically. But we need, I suggest, to think beyond the genome sequence as the only way to measure risk. The most important way of measuring risk is to ask what is the product of the genes, and that is going to be seen in the proteins that are made by the genes and by the metabolites that are the products

of the way that the proteins transform energy in to energy out and build up the cells, and there are different ways of developing those technologies which are starting to get the same kind of push that the genome technology favours in the last 20 years or so. That is going to be much more important because that is closer to the actual output, the phenotype, the clinical measures that we are interested in, than simply having everything measured at the level of the gene. So I think that we will see technologies which will replace the kind of glucose dipstick which are now profiling all the products that are relevant to the determination of your glucose level or your cholesterol level and so on, and that will be based on evidence that we glean from biobanks and genomic medicine studies.

**Professor Morris:** Just to add to that, I think the vision is one of personalised medicine. One could argue that if we give ten people a drug today in three people it works, in three it does not work, three people do not take it and one has an adverse event, and that is the current equation for commonly used therapies. So I think pharmacogenetics is part of the vision of what these large studies will deliver on so that we can actually predict reliably therapeutic response based upon a baseline genetic profile. I think in terms of the clinical impact of that it is an important part of pharmacogenetics as well as clinical risk prediction earlier on in the life course of disease progression; so that at an earlier stage we can make predictions of risk so that either lifestyle or therapeutic recommendations can be made earlier on in the life cycle.

**Professor Sampson:** If I could answer your question from a slightly different angle and that is I was thinking of the utilisation of these technologies in the clinic in the future. There is a lot that is not known about how that would be implemented in terms of professional understanding and also public understanding and uptake, and so in parallel to these investments in scientific advances it is also very important that social science studies are undertaken and that adequate investment is made in education and public engagement. There are a number of organisations within the genomic medicine community in the UK that are

involved in this aspect of the work, including ESRC funded genomics network with centres, including the centre in Cardiff and the centre in Edinburgh, amongst others; and also the gene knowledge parks which were initiated as part of the government's White Paper on genetics, which is still taking a major role in relation to educational initiatives.

**Q498 Lord Colwyn:** Do you believe that the translation of the records that you get at Biobank are 100 per cent accurate because we are already hearing about private companies which are analysing the same samples and getting different results?

**Professor Collins:** I am sorry, the translation of which records?

**Q499 Lord Colwyn:** Of the genomic information leading to susceptibility of disease. Are you happy that you have it right because we are hearing other companies which are taking the same sample from the same person – three different companies – getting different results?

**Professor Collins:** Yes, there was a nice example of that in, I think, the *Guardian*, where a journalist compared the prediction of risk by deCode – 23 and Me and another company. Personally I think that the approach of these companies is simplistic. They are based on looking at genetic variants in isolation; they are based on studies which are very small where you can possibly pick up the effect of a variant but you do not how it interacts with other things, both other genetic variants but also environmental or lifestyle factors. So I am actually not surprised at all that they can come up with different results because I think the data on which they are based is weak. That is why we need much bigger resources that will allow us to look at the relationship between a range of different risk factors in order to get a complete picture of a prediction of disease. But I do not think it is only about the prediction of disease. For example, we may well find genetic variants that produce only very small effects on risk, but what that could mean is that we have identified a new pathway for disease. That pathway could then open up the discovery of treatments that would act on that pathway

that could be of substantial benefit. An example is the statin drugs which are used in millions of people around the UK and tens of millions around the world and saving tens of thousands of lives each year. The statin drugs is a nice piece of observational epidemiology like UK Biobank, which shows that cholesterol is an important risk factor for disease and genetic work that shows that there is an abnormality in the receptor – the LDL cholesterol receptor – that influences the amount of bad cholesterol, the LDL cholesterol that you have in the blood. The statin drugs interfere with the pathways that can affect that LDL cholesterol receptor. They can make someone who is abnormal genetically normal and lower their cholesterol and lower their risk and it is a combination, an understanding – I think this is the key thing – a knowledge of disease process from epidemiology and from genetics that has led in that case to drugs that are saving tens of thousands of lives each year.

**Q500 Lord Broers:** Can we backtrack to a simple matter of sequencing. The data we are hearing – at least that I am hearing – are not consistent. There is a US website offering to sequence your DNA for \$999.90; is that a complete sequencing? And is complete sequencing required? We have also learned in our papers that you can sequence the entire coding regions of BRCA 1 and 2 for perhaps £10 soon. So could you talk about that? Do we need a complete sequence or is much of the sequence irrelevant? And what is the price anyway?

**Professor Porteous:** The price is tumbling – we have said that and I think we need to restate that. In a sense that is less important than why you are sequencing and what you are looking for. I would just like to echo Rory's comments by emphasising the value that perhaps we have not yet teased out yet adequately about the biobanks. You have heard in your evidence previously about the really remarkable success of the Wellcome Trust case control consortia, where large groups of investigators have come together, amalgamated their samples – collected some 2000 or so cases of Type 2 diabetes or Crohn's disease or one or other disease – then scanned the whole genome and looked for signatures from genes that suggest that there

might be a risk factor there. That has worked extremely well but it still only captured between ten and 20 per cent of the total genetic liability that we know lies somewhere in the genome that has yet to be found. Therefore we have valuable but still partial information. So we need to do two things; we need to look for the remaining information, where it comes from, and we need to seek evidence of whether or not those findings made in one study are replicated in a second, independent study. Moreover – and this is where I think the biobanks come into play – it is very important that you validate those studies in the general population rather than simply in populations that have been acquired because they are a case of inflammatory bowel disease or a case of diabetes. What we want to know is how much do those genetic factors play in the risk of an individual later in life developing diabetes or later in life developing Crohn’s disease or depression or any of the other psychiatric disorders that you have also taken evidence on. So I think that there is an attempt by a number of speculative companies to get rich quick. There will be people gullible enough to take up their offers of genetic testing, but I would have said that the chance of the results being meaningful is extremely low because the evidence base is far from complete.

**Q501 Baroness O’Neill of Bengarve:** I would like to come on to the question of the information that you need. We have heard several times about differences between the way that centralised healthcare records are maintained in England, Scotland and Wales – so you are the perfect group of people to enlighten us on that. We have also heard that the healthcare informatics framework in Scotland works better as a basis for research than that in England. Do you think that that is the case and what are the differences and what makes one system better or worse than another for research purposes? A large set of questions there.

**Professor Morris:** Shall I kick off as it is my area of interest? An initial comment would be that we should never ignore the phenotype because there is so much emphasis placed on genotype; but arguably, I would say, that phenotype is actually of equal if not more

importance. An example is that Rory says that UK Biobank is powered by incident or new cases of disease. What we absolutely require is robust mechanisms to ensure that we pick up all new cases of disease in these very large cohort studies. To do that efficiently I think we have to look at the domain of health and informatics to support these research programmes because otherwise if that efficiency is not built into the system the costs associated with these studies will spiral. In terms of Scotland, if you look globally the healthcare economies that do this well have a certain size – about five million people – so I think that size is in Scotland’s favour because it is much easier to do with a population denominator of five million people; allied to a low rate of migration, allied to a single health provider, NHS Scotland, with less penetrance of the private sector; also allied to a collaborative inter-disciplinary approach to answering these questions. So on that sub straight I think that Scotland arguably is ahead of the game because since the 1970s we have used a single patient identifier for every episode of healthcare. This is not rocket science, but in general practice we have used the so-called community health index number to track all episodes of healthcare, and this is now being implemented throughout the health service and is reaching 93 per cent implementation in hospitals. I think that single fact has given us a competitive advantage; it allows us to track people through the fragmented journey of care from their general practice or laboratory to the pharmacist and the hospital. What is the research benefit? I think a great example – again going back to statin drugs, one of the biggest studies of statin drugs was led by Ian Ford in the west of Scotland and he in the mid-eighties tracked people in the context of a clinical trial for four years, and he showed that taking a statin reduced the risk of heart disease and stroke. That study probably cost £25 million to £30 million. What he published in the *New England Journal* at the tail end of last year was the ability to track these people; they were no longer part of the study but he could track them for vital events by linking into hospital admissions, and he published the 15-year follow up of the initial study. The key here was that the cost of

the 15-year follow-up study was only £60,000 because of the efficiency of the linkage into vital health events. What he showed was two things: firstly, the protracted and prolonged benefit of statins in terms of cardiovascular risk; secondly, he put to bed the suggestion that statin use was associated with cancer development – so they are safe. So drug safety and drug efficacy. Scotland, because of its integrated records, is able to run such studies. It is not without its problems but perhaps I will let others comment before I come back to the problems.

**Q502 Chairman:** Is there a comment from either the Wales or England side?

**Professor Sampson:** I can comment a little on the situation in Wales. In Wales there is a developing centralised NHS healthcare informatics system and this is termed Informing Healthcare. It is certainly behind the state of development of the Scottish system. It was initiated in 2003. It is a single system across Wales and it is being developed in Llenelli, so it is done by evolution rather than revolution. In addition to that NHS-based health informatics system, there is in parallel a research system which is under development at HIRU, which is the Health Information Research Unit in Swansea University. That unit is developing algorithms to assimilate anonymised data for research purposes from the various health databases that exist in primary and secondary care. At the moment neither of these – the health informatics systems, even the identifiable but encrypted NHS one nor the anonymised research system – are actually linked into genetic data and there are concerns about the confidentiality of genetic data in relation to these systems. But there certainly would be potential for these integrated clinical health information systems to be linked to genetic data in those circumstances where consent has been given, for example in Biobank. Indeed, there are significant developments now linking these health systems with the clinical health information systems through to, for instance, the Wales Cancer Bank, which again is a consented biobank.

**Q503 Baroness O'Neill of Bengarve:** Could I just ask for clarification whether you mean by anonymised they are irreversibly anonymised and de-linked or whether you mean that it is anonymised but a key is held.

*Professor Sampson:* A key is held but it is double encrypted.

**Q504 Lord Warner:** I am slightly puzzled by some of this – and I say this as the guy who had responsibility for Connecting for Health in England as a minister. The custodians of these records are essentially clinicians and the biggest obstacle to integrated use in England was the clinicians, who are deeply suspicious of integration and who are very protective about individual records. Is there something different about the clinicians in Scotland and Wales? Is it something about population size which actually makes the clinicians behave differently because they have been the biggest resistance or integration of records in England?

*Professor Sampson:* I think there probably is a difference; it is a smaller community. There is, I think, a bigger sense of collective ownership and perhaps a stronger socialist tradition in Wales, and I think there are cultural differences which may make these easier as well just the logistics of scale, which I think is a very big factor.

*Professor Morris:* I think you have put your finger on a key issue. The databases I alluded to are the national NHS Scotland administrative datasets and they may be rich in clinical content – for example we have the national screening systems, we have the national hospital admissions database, outpatients' database, GRO, and they are coded according to the international classification of Disease 10 or operating codes. So they are rich in clinical functionality, yet they are administrative databases hosted by NHS Scotland.

**Q505 Lord Warner:** So you bypass the clinician cooperation issue in a sense because they are there.

**Professor Morris:** I think “bypass” perhaps is a strong word! The governance of the system includes clinicians but the data sources subject to good governance are under the aegis of NHS Scotland. Just to follow up, the real prize for the future is the linkage of clinical datasets, as you say, and that is where the challenge lies because we, for example, our healthcare economy, we have a budget of a billion pounds a year. We probably have about 150 clinical systems, each with their own jurisdiction, and if we are really going to yield the value of these large cohorts we actually need to get into the granular detail of these clinical information sets. I think that is where the challenge lies because of these issues of clarity of governance, clarity of data protection, clarity around confidentiality and there is a lot of work that needs to be done in that domain if we are really going to see the value from clinical datasets.

**Q506 Baroness O’Neill of Bengarve:** I think we would like to hear from Professor Collins about England in this respect. How does it look for Biobank UK? I should say that I enrolled and I know the process and the consent seems to be fine, but of course you have to link to subsequently developing clinical information. Is Connecting for Health to provide that? How is it to work?

**Professor Collins:** I think personally that this is the most critical issue for many of these studies. It is not the technology – that is developing. It is building these resources and making them as informative as possible. If you look back at the studies that have been done what they were able to do was to look essentially at the diseases that killed you and at cancers because there were registries of death and cancer. Why did Richard Doll do the study in British doctors? Because British doctors got paid if they were registered, so he could follow them. It is the ability to follow people and to find out what happens to them that allows you to really increase the value of these studies. The issue that we have is not being able to follow the diseases that do not kill, but that disable, that maim, that cause misery. Dementia, joint

disease, why are these not as well studied as cancers and heart disease? It is not because they are less common, it is because we cannot follow them in many of these studies. So the ability to link participants in these studies to their full health record is absolutely critical. Scotland is great but we are seeing increasingly the ability in principle to link to more and more records. So, for example, not just death from cancer but in England to hospital statistics so that you can find out at least the diseases that get you into hospital. In principle you could link to pharmacy record details so you could find out what drugs people are getting and look at pharma vigilance in terms of safety of drugs. I do not think the problem is that the in principle ability to link, it is actually more a problem about being allowed to link. It is the bureaucratic obstacles to linkage that are the concerns. They were highlighted in the Academy of Medical Sciences' report on personal data. They have been, I think, understood by people in Connecting for Health by developing the Research Capability Programme, which is looking at ways to facilitate research. But actually what is required is a desire, a willingness to understand that research is at the heart of healthcare. If we do not have research then we do not good healthcare, and it needs to be seen as part of the healthcare system.

**Q507 Baroness O'Neill of Bengarve:** Just one very small point there. Do you think that it is fundamentally a bureaucratic or a legislative problem? Is it the data protection legislation with its many intricacies, if not incoherences, or is it the administrative framework that has grown up around its attempted implementation?

**Professor Collins:** It is both. This is the summary of this report by the Academy of Medical Sciences on the working group. The legislation is not clear; it can be interpreted in a variety of different ways, apparently, and the consequence of that is that different people with different agendas interpret it in different ways. So I think it is a combination of legislation which is unclear and therefore leading to differences of opinion.

**Professor Morris:** Can I build on that because I think this is an absolutely pivotal issue? The Department of Health guidance suggests that this domain is affected by 43 relevant pieces of legislation. There were 12 sets of relevant standards and eight professional codes of conduct and what it has bred is a culture of caution, confusion, uncertainty and inconsistency. Even Lord Phillips suggested that for Lord Phillips it was like weaving his way through a thicket. So for us to interpret it and to have consistent interpretation from legal bodies who have data protection responsibilities is absolutely key. Currently this is a if not *the* major issue in terms of the ability to safely link linkage which is in the public good with appropriate security, safeguards and orders. I think that has been a major focus of the Walport Report, which was very welcome.

**Q508 Baroness Perry of Southwark:** I am puzzled as to how you are going to get the kind of follow-up data that you want. Is it not your intention to keep in touch with the people who were in your cohort and revisit them every seven years so that you do keep a record? And if you do not have that facility then collecting the biobank is going to be severely damaged in its usefulness, is it not? If you are looking for the sample of your subset that develop heart disease or whatever you are looking for, how are you going to find them unless you keep bringing the subject back? I speak as an enrolled person in the Whitehall Study, the stress and health study, which has been going for over 20 years, and we are brought back every three years or so to find out which of us have died off or gone off in the meantime. That is the essential part of the value of the study, is it not?

**Professor Morris:** With respect to UK Biobank certainly one needs to keep in touch with disciplines but the thing that one is interested in is what has happened to them primarily in terms of health outcome, so that linkage can actually be done in principle through health record systems. As I have mentioned, for the whole of the UK, death, cancer, hospital episodes can be linked. Already in Scotland one can link beyond that to primary care.

**Q509 Chairman:** To clarify, how does the UK Biobank currently link with databases, GPs, hospital episodes and death records?

**Professor Collins:** At the moment we are not linking but we have approval from what was the ONS to link to death and cancer registration and you can then link to the HES system – Hospital Episode Statistics System. We have piloted that in collaboration with Professor Burrell in Oxford, who has linked the 1.4 million women in the Million Women Study to the Hospital Episode Statistics data. As the ability to link to more record systems, with the consent of the participants which we have, becomes easier, then we will be able to do it beyond Scotland (where you can already do that in primary care) to those other record linkage systems. In terms of seeing participants again, we will be seeing a representative sub-sample of the participants in UK Biobank, not to find out what happened to them in terms of health but in order to assess the variation, the changes over time, the risk factors in the way they live which will inform us about the whole cohort. To be efficient, we do not have to see all half a million again; we will see 20,000 or 30,000 every few years in order to track variation in the ways in which they live. Obviously their genes will stay the same, but it is these linkage systems which allow us to look at their health outcomes. I should also add that the linkage systems can allow us to look at health-related records systems. We can look back in time, at their health before they went into the study and at other things, like occupation, which will tell us about risk factors, exposures, as well as looking at health in the longer term.

**Q510 Baroness Perry of Southwark:** If you are only looking at the health records, the hospital records and so on, and the records of death, are you not missing out on the people who may have the same genetic variation but who do not get that particular illness? Is it not just as important to know that 10,000 people who have a genetic variation all developed bowel cancer or something, but another 10,000 or 5000 or so who had the same genetic

variation did not. Is that not just as important? You will not take that up from the health records, will you?

**Professor Collins:** Yes, because we are linking to the records, and the people who have the disease would be in there and the people in the Biobank project who do not get the disease will be linked as not having had that condition at least diagnosed, and we then compare the people with and without the diagnosis.

**Q511 Lord Warner:** I am trying to understand this. This is quite a critical issue. Are you saying that all the people who are in Biobank have given consent to you going in, through the connecting for health, presumably information through the spine, from the GPs to give them the information that you request from the GP? I am trying to understand this, because this seems to be quite a critical issue.

**Professor Collins:** We could ask one of the participants what they felt they had consented to, but I will answer, first, for them. Yes, there is explicit consent from participants for linkage to health-related records. That is defined as not just medical records, but other records – and it gives examples – including, for example, employment records both in the past (that is, before they went into the study, so that tells us about exposures before they joined) and also their health-related records in the long term through their medical records systems.

**Baroness O'Neill of Bengarve:** If I remember – it is quite recent, and I should – there are ways in which, as you consent, you can exclude certain types of things like attendance at STD clinics and perhaps abortions – I cannot remember other things but there were things like that – but otherwise it is consent. I would add that I happily gave my consent to the lot. Having always been treated on the basis of knowledge that was obtained by treating other people, I do not feel this information, if duly anonymised, is in any way private or mine. But I may be one of the few people who thinks that. The law does not.

**Q512 Lord Warner:** We have had a good go over, in a sense, whether the present IT systems are sufficient for joining up information from different databases. I think the sense in the answers we have had so far is that there is still some way to go. Could we explore a bit more how we might improve the joining up because it is not just about technology? Can we explore some of the other issues that you think need to be tackled if we are to join these databases up?

**Professor Morris:** We have touched upon the clarity of governance and the legal status of linkage. Currently the situation is that the clinical data is very much silo'd and it sits under different jurisdictions. I think, therefore, we need to look at the recommendations of the Walport Review and come up with a response to that which is clearly in the public good as good governance, and legislate if required, because there are differences between England and Scotland. Currently, I understand that PIAC have legal status; whereas, in Scotland, the Privacy Advisory Committee of ISD which oversees the governance of these national datasets is advisory and has no legal status. I think we need clarity around that. Second, I think we need to get on the front foot in terms of public and professional education. This is something which the Information Commissioner is very strong on. He says we should be much more forthright, informing people about how we use data, why we use data, the benefits of using data, and the safeguards which are in the system to ensure confidentiality and security. Certainly in Scotland, in 2002 a very good report, called the CSAG report, recommended a public consultation campaign to demonstrate the clear value that research has in informing and improving health services. That campaign has never happened, so I feel there is work to be done about informing the public about how we use information and the benefits of using information. The third issue is around trusted models of linkage. Currently we have these multiple data silos with huge variance in the way that data are linked or what is allowed to be linked. I think we should look beyond health at other examples of things that work. A good

example that I think we need to look at is work which has come out of the ONS called the Virtual Microdata Laboratory. They have built a very resilient and secure system, with “thin client” secure and safe and audited access from multiple sites from the UK. They have been able to link very confidential person-specific datasets for the benefits of research. Researchers go into a protected environment. They are protected, they are overseen, but then they are given “thin client” access, which means that they are not allowed to download any data. They are given access to a central data repository for research purposes. The virtual microdata laboratory is seen as a huge success. It is now a model which is being adopted throughout Europe. I think there are three broad areas where we need to focus our attention so that this linkage is seen to be in the public good and also has industrial strength in terms of its security and confidentiality.

**Q513 Chairman:** In the long term, all these databases will need to be linked to genetic databases.

**Professor Morris:** There are various ways of approaching that. On a case-by-case basis, yes. Currently, the way that we think genetic data is that, following ethical committee approval, Caldecote approval, NHS R&D approval, we are allowed to link for specific research questions and that process has a very distinct audit trail.

**Professor Porteous:** Clearly, the discussion we have had at the moment has been about linking health records to genomic data, but we should not lose sight of the fact that we need also to have IT solutions to linking and securing and auditing the genetic data. It should not be taken as a simple “press of a button” and you get an accurate answer out with certain technologies. You have to have laboratory information management systems in place – which we have in Generation Scotland and which I know the UK Biobank has – that can accurately track and audit all the samples that are received, and do that linkage too. If you have inaccuracies, is always the weakest link that is going to be where the problems arise. It has to

be said that much research in the past that has been done in this area has been done very much on a short-term basis. I think one has to be cautious about, for instance, using legacy data (samples have been collected in the past, stored in some clinical researcher's fridge or laboratory and then taken out subsequently), to be sure that you are matching up with modern technology to material that was stored in the past without those kinds of secure ways of coding the information, coding the samples, and then pulling it all together. There is a laboratory side to this IT technology that has to be addressed as well. But that is easier to do, because you have control over it, but you have to be very careful when you pull that data together from different biobanks to make sure that all the biobank information you are getting from around the world is set at an appropriate standard. That is where organisations like P3G and HUGO and others come into play to try to set those international standards.

**Q514 Lord Warner:** We have spent a lot of time painting a picture of a world in which there is better linkage between big databases. There is a burgeoning volume of information are available. Do you have any concerns about whether the informatics skills capability is going to be up to the job? What do we need to do if it is to secure benefits from this kind of linking of the huge amount of information?

**Professor Porteous:** I think this is a very exciting opportunity. The physicists who have developed IT solutions for large number calculations have skills that we are barely aware of. I can give you a good example using a problem that we had. We were frustrated by the inability of a well-used statistical algorithm that is used to try to glean information out of genetic data: it is well-used in a certain setting, but when you try to use it to look at larger segments of the genome, it just falls flat, it cannot do it. We took the problem to our local colleagues at the National e-Science Centre in Edinburgh, and they developed, within a matter of a month, a grid-based solution to this that cut the computational time a thousand fold and gave us a working solution. At that level, I do not think we have a problem. The only

problem will be if we think that our home-made solutions are adequate. We need to talk to people who are real experts in the field at that level. They are there; they just need to be alerted to the opportunity and the need that we have and I think they will be able to respond.

**Professor Collins:** I want to agree with that. I think that we are generating resources that will generate huge amounts of data – not just the genetic data, but the protein metabonomic data. It will be working out what is real in those data that will be the most difficult part of it; it will not be the technology of measuring it. Encouraging people with bio-informatic skills to work in the biology area is a critical one, and building a career structure for those individuals. The Wellcome Trust had a meeting in January at which they discussed this. One of the key things which came out of that was that we need to stop thinking about these people as information technologists and start thinking about them as information scientists, because they will be the ones who help us to tease out what is real in these resources. Maybe we should see the banking crisis as an advantage, because we can recruit all these people into biology. Once they come in, they will not go back because it is a lot more fun!

**Q515 Lord Warner:** If a bit less well-paid.

**Professor Porteous:** A steady, low income

**Q516 Baroness Perry of Southwark:** One of the issues we have constantly talked about and had evidence about in the field is what you tell the subject of the investigation after they have given their samples. Particularly as knowledge advances, it will become apparent that some people have a very high risk of a disease which may be curable – and that is one reason for wondering whether to tell them – or may be incurable – and that raises different ethical issues. Are there any plans, if these do emerge, to inform the subjects and their families?

**Professor Collins:** I could speak for UK Biobank. This was discussed at great length and included wide consultation with patient groups and the general public, as well as ethicists,

lawyers, doctors, and researchers. The conclusion we came to – and this was really signed off by the Independent Ethics and Governance Council for UK Biobank – was that no feedback of individual results would be provided to participants. The invitation to take part and the consent is explicit in saying that people understand there would be no feedback of their individual results. There are a number of reasons for that. Partly, we are not doing any assays of the samples upfront (we are doing them in five, ten, 15 years) and we do not know what we are going to measure, so it is impossible to counsel people on what the implications will be of things we might do in 15 or 20 years time, or, indeed, what the relevance of the things that we find is. Going back to previous comments about these websites where you can find out your risk and it is widely divergent, we do not yet know when we get research results what their full impact is on disease, and we do not know everything about that patient in terms of their health like their own doctor does. We are feeding back the findings of the research that is done on the study. That will be made available to all the participants and to other people as to what has emerged. If findings emerge that should change health care, in that it might include screening tests that are validated and really strongly predictive of disease, then those should become available in an appropriate way through normal healthcare systems. But we get explicit consent that there will not be individual feedback of results.

**Q517 Baroness Perry of Southwark:** Do you see this as being permanent, or might it change in ten or 15 years, when both the technology has changed and the findings that will emerge over that time may make prediction much more accurate?

**Professor Collins:** The idea that the fund has had of setting up the independent Ethics and Governance Council for UK Biobank was that things do change. Their role is an oversight role and to provide guidance. I cannot say that things will not change. They are there to guide us on that. Participants also have the opportunity to voice their concerns, because they can withdraw at any time. They can let us know if they have concerns about the ways in

which the resources are used. There was really very detailed discussion and consultation for UK Biobank about the pros and cons of feedback. I cannot see the rationale on which that was based changing, but I cannot predict the future.

**Q518 Lord Colwyn:** We have covered some of the problems of incidental findings, that are already causing, for instance, dilemmas such as this discovery of abnormalities in brain scans. Genome studies are likely to accumulate much more information and have an effect on health, so privacy and confidentiality are very important. What are the risks and how are you managing the risks to privacy and confidentiality for participants and their families in biobank collections? What role do patients and volunteers play in deciding what will happen to their sample and information supplied as part of biobank studies?

**Professor Porteous:** Perhaps I could reflect on the previous question first, because I think it does follow on from the one you have raised. The situation with Generation Scotland is very similar. There was a broad view taken and we have very much followed and, indeed, been part of the discussions that led to the formation of the UK Biobank. Generation Scotland has taken a very similar view and operates in a very similar principle; that is, if something clear and important and overt of a clinical nature emerges at the point of the clinical examination, then that information is fed back, but thereafter not. We are very keen that we emphasise the fact that we are doing studies that are looking for predisposing factors. To go back to the points made earlier by Andrew, the process of taking that evidence through to clinical application is one that requires replication and validation and several steps in that process before one would have, for example, a health technology assessment view as to whether or not something had diagnostic value and should be implemented. For all of those reasons, the notion of giving feedback on an individual basis we think is inappropriate; but like UK Biobank, we too have an advisory board, set up at arm's length from Generation Scotland by the Chief Scientist Office, chaired by Lord Sutherland, with advisers who also link between

the UK Biobank and Generation Scotland. Graham Laurie is a member of both. Thus we have, I think, a good way in which we can see at a UK level, both UK Biobank and Generation Scotland, issues as they emerge and to get good advice from external groups as to how those should be handled. So far as risks to privacy and confidentiality are concerned, that is clearly of great import, because without that confidence in Generation Scotland or UK Biobank, we will neither gain nor retain participants. Of course, the retention of goodwill is particularly important for a family-based study, so we have been given this great thought and, we believe, care and attention, to ensure that we build in, through the IT structures, the appropriate firewalls that prevent information flow in anything other than a fully controlled fashion. I am sure Andrew will go into that in more detail as to how that works. I think you are right to raise the question, but it is one that we have spent a lot of time deliberating on. But, particularly for a family-based study, it is paramount that those entering the study feel confident that they wish to be part of it, feel confident in the reason for the research being done, and feel confident that the measures that we have in place secure that information and that it will be used for approved research only.

**Q519 Chairman:** Professor Collins, would that apply also to Biobank?

**Professor Collins:** Yes, we have the same measures. There are two parts to your question. The first part relates to key code anonymising data, so separating identifiers from the data and keeping the identifiers separately and the codes separately and secure; encrypting data, particularly when it has been transferred. We spent about a year getting permission to get names and addresses to invite people into the study from the NHS, and it took us a little bit longer than to persuade the NHS to stop sending them in unencrypted data files. So it is doing things which make sense: as I say, protecting the identifiers and keeping them separate; having firewalls; paying people to try to hack into your system – we pay the NCC, a national organisation, to try to hack into our systems and identify potential areas of failure; and only

allowing named individuals, a limited number of individuals, to have access to identifiers. In addition – and I think it goes to one of the concerns that was raised in your question about being able to identify people from genetic information – in terms of what researchers will have access to, it will be anonymised or key code anonymised, but they will also have to sign agreements that they will not use the data to try to identify individuals, so there will be contracts with researchers in that respect. On the second part of your question: What control do participants have over the users? they can withdraw at any time. We tell them about the records we are linking to, we tell them about the uses of the resource, and if they feel that there is a problem with that, then they have the right to withdraw at any time and pull everything out of the study. They also have the Ethics and Governance Council to raise their concerns with. So they really have control.

**Q520 Lord Colwyn:** If the information is left in or stolen from a train or a car, it is of no significance.

**Professor Collins:** Yes, because, for example, in our assessment centres, the participants enter data directly onto computer using touch screens and those data are encrypted, so if someone comes and steals the computer it is encrypted data and it cannot be broken into.

**Q521 Baroness Perry of Southwark:** I think, possibly, your previous answer has answered this question, but we are told that there is recent evidence suggesting that an individual can be identified from their genomic profile even if it is present only in summary format amongst hundreds of others. What are the real risks in publishing this data in medical research or when linking different databases? Perhaps, at the same time, you could answer this question: We understand that while DNA theft is a crime, taking a DNA sequence maliciously to identify someone is not a crime. Should it be? Have you had any consideration about that, because the pressure to criminalise it would come, I suppose, from people like yourself?

**Professor Collins:** In terms of the use of the genetic data which researchers would need to be able to see, they will only be able to do it under contract, where they will be in agreement not to identify individuals. If they tried to identify an individual, they would be in breach of that contract. Recently there has been a publication saying that in summary genetic data you can identify whether an individual is in that summary data file. The consequence of that has been that, whereas, for example, journals such as the New England Journal of Medicine have required you to put the summary data on the data repositories, they are stopping doing that because you could, in principle, identify individuals. We in UK Biobank are not going to be putting such data out into the public domain: it will be available only to researchers under strict control. I do not see it as an issue, therefore, from that point of view.

**Professor Porteous:** The point that has been made is that you can identify that an individual is in a group rather than identify the individual. It is “which sample corresponds to a particular individual” – which is, I think, rather different. One of the things I am often struck about is the way in which genetic studies highlight issues which are out there in other guises. We assimilate quite readily, whether it is Facebook or Myspace or other ways of identifying individuals and collecting an awful lot of highly personal information from them. I do not want to use that in any way to minimise the potential concern, but I think we should try to make sure that we are clear about what exactly it is we would be revealing and how it would be used. In terms of a governance issue, exactly the same rule applies for Generation Scotland as do to UK Biobank: you would be in breach of contract if you were to use such information in that way, because, again, the terms of access and the use that is made of the samples to do research would preclude the identification of individuals. I think it is a theoretical issue, but it is one which is addressed in many, many other ways that does not include genetic profiling. It certainly is true in terms of DNA fingerprinting that one is able to identify individuals through genetic analysis, but that is done using a rather different type of

genetic information and it is done with a very express intent of identifying an individual and distinguishing one individual from another. Around that, we have a very clear set of frameworks that determine when you may and when you may not use forensic fingerprinting technologies.

**Q522 Baroness Perry of Southwark:** What about the issue of making it a criminal offence maliciously to do it?

*Professor Porteous:* That would be captured under the existing legislation, I believe, in the sense of DNA fingerprinting technology. But you are making a distinction between the theft of the data as opposed to the theft of the sample?

**Q523 Baroness Perry of Southwark:** That is right.

*Professor Porteous:* That would be interesting. I am sure that would generate a wonderful legal debate. One could, I am sure, argue that you had technically stolen the sample in order to generate the data, but I am not a lawyer, so I am going to be quiet now.

**Q524 Chairman:** Earlier on, Professor Morris, you mentioned that a public education campaign was required, so that the public could begin to understand the implications of the vast amount of information that could become available in two, three, four or five years. Do you think that is so? How do you think this should be handled? Equally, is there a necessity for the education of professionals in genetics, in association with genetic information and health information or risks related to that?

*Professor Morris:* We touched earlier on the issue around information scientists. The one key message I have learned from biobanking is its interdisciplinarity. If we are going to get this right, we need to bring disciplines together. I am pleased that we have attempted to do that with Generation Scotland from day one. We have brought together information

scientists, statistical geneticists, laboratory scientists, and doctors, but one of the key groups has been the social scientists. From day one they have led a research programme which has actually informed the development of Generation Scotland and its operation. They have run focus groups with families, excellent questionnaires, a Mori poll, and it has shaped the protocol and our approach to communication with the public. And they have looked at thorny issues, in terms of pharmaceutical access to data. They have looked at the issues about the right to withdraw our policy on access to the police and they have also commented on benefit sharing. That research programme from day one has been critical and it has given us the confidence to move at a pace, so that we do not move forward in a vacuum. I think that is the message. Another part of Generation Scotland has been to work with our clinical genetic colleagues around professional education. Not only in their community but also in the great community we have of primary care, this science will, we predict, be impacting upon their clinical decision-making within the five- to ten-year time frame and we need to prepare that ground so that they are ready for it. As part of the ScotGEN network we have developed educational tools and website educational programmes to try to educate the clinical community about the impact that this genomic revolution will have. I think it would be folly to wait until the yield of these biobanks is realised before we tackle these issues.

**Q525 Chairman:** What about England and Wales?

**Professor Collins:** To echo a previous point, a lot of the things that are discussed about genes are not new. Particularly in terms of risk factors for disease, we need to think of them as one part of it. It is the genes, the gene product, the environment, the lifestyle, and, in a way, there is a little bit of hysteria around genetics, as if they program you for disaster. I think we need to get across the message that they are one part of a number of things that can determine your outcome. The advantage of many of those other things is that you can change them. That is one key message that one needs to get across. I believe that participation,

particularly on a population-wide basis in such states, is a great way of understanding what they can produce, in being part of something that will improve health. I can give you an example of this. The Wellcome Trust funded ALSPAC study is a study of children and mothers in Bristol. We are now recruiting in biobank in Bristol and our recruitment rate there is twice as high as in any other part of the country in which we are participating. I cannot tell you that is because of ALSPAC, but many people who come in talk about the ALSPAC study. It is part of their commitment to help research. I think these population studies, like Generation Scotland and like UK Biobank being part of how the population helps to improve their health, is one very good way of getting the population to understand the benefits of medical research, and for their doctors, also, the general practitioners particularly, to understand the benefits of participating in and helping such research. I think the studies themselves are educational in that respect.

**Professor Sampson:** I would concur with what my colleagues have had to say about the opportunity that these large scale studies have for engaging the public in a wider way in relation to their benefits in medical research. It is not an explicit role of these biobank projects, but it is an explicit role of a number of organisations. Earlier I mentioned the Gene Knowledge Parks, which were initiated in response to *Our Inheritance, Our Future* White Paper from the Government. Those organisations have done a lot to bridge the understanding gap that exists between scientists and healthcare professionals and the general public in relation to genetics, and to try to address this perception of some hysteria or suspicion of genetics exceptionalism that has been, perhaps, prevalent in the public. I think a number of these initiatives have shown that the public are very receptive to and very tolerant of the exploration of genetic issues and genetic research. In Wales we have undertaken a number of public engagement programmes that have involved thousands of young people, in particular, but also members of the wider public. For example, a citizen's jury on designer babies

showed that young people are very liberal in their approaches and their attitudes to the application of genetics in clinical and healthcare settings. We have also undertaken a study in relation to the national DNA database, and we are now undertaking a further study which is engaging with the individuals who are on the national DNA database to explore their perceptions of the ways in which genetic data is used in the legal setting. There are a number of organisations and initiatives around the UK that are doing a lot to address this gap in understanding between the scientists and technologies (that are moving at a heck of a rate, as we have understood) and the general public. But the message is that attitudes, particularly amongst the young, are optimistic and quite liberal.

**Q526 Chairman:** Could you give a brief the answer to a question which relates to some of the information in evidence that we were given, that a bigger European wide Biobank would be of great value. Would you comment about that?

**Professor Porteous:** Certainly there is scope for funding this type of work under the EU Framework 7, and it is under discussion for the next round of package for biomedical research support. There are, indeed, a number of studies which are already funded under through that programme but they tend to be of a more modest nature and tend to be more disease specific, disease focused. But I think there is a considerable amount of interest from mainland Europe to what is going on in the UK and a strong desire, if not to directly collaborate, then to try to find ways in which one can capitalise on the work that is already going on in the UK and see it replicated around Europe. There are, for instance, studies such as Genome EU Twin, which has a twin-based study that captures information through twin studies all over Europe. That is also paralleling the kind of work that we do in the biobanks. There are other technology programmes, such as MoIPAGE, which is developing the next generation of techniques to make measurements of biomarkers and they will feed in. Whether it would be a good investment to have the UK Biobank equivalent in each of the EU Member States, I think

would probably take a considerable discussion and debate, but it would be up to each of the Member States to decide whether or not they thought it was a good investment and value for money.

**Professor Collins:** I think there is merit in having such studies in a wide range of circumstances. I mentioned a similar study in China, where we can look at leaner individuals, we can look at people with lower cholesterols. There are studies going on in populations which are perhaps not where we are coming from but where we are going to, such as studies in Mexico where there is much more obesity and much more diabetes, so we can look at the other end of the extreme. Having studied in different circumstances allows us to look at a wider range of exposures. In terms of investment in Europe, personally I think that phenotyping is the key to these studies and more detailed information about the participants in terms of their risk factors beyond genetics will increase their value. Although I am very keen on bigger studies, I think there is merit, also, in taking the big studies that we have established an, increasing their value by increasing the detail of phenotyping.

**Q527 Chairman:** Finally, if I were, without any promises, to ask you which two key recommendations you would like to see, what would they be?

**Professor Collins:** If I could have one, it would be to remove the bureaucratic obstacles to using health records to improve the health of people in the UK. We are uniquely placed here to do these kinds of studies, to do large-scale clinical trials. The regulatory burden on the use of records and the regulatory burden for clinical trials as a consequence of the EU Directive on clinical trials and its implementation into UK law, have pushed research and research funding out of the UK. I think the big problem concerns the regulatory burden and the bureaucratic obstacles. The consequence of those and NHS research governance is making research increasingly difficult and our ability to do this kind of research increasingly difficult and costly, and is slowing it up substantially. So I only want one thing.

**Professor Sampson:** There is one stage back from that even, which is perhaps the balance between individual rights and responsibilities as part of society. A lot of what we are talking about is being driven by the changes in that balance towards trying to protect the individuals at the cost of harming the greater good.

**Professor Morris:** I would suggest that the UK could be at the forefront of translational medicine internationally, not only going from bench to bedside but what we have talked about today, which is from cell to community. To do that, I would like to endorse Rory's point. I think that it is absolutely key: to have clarity about use of data, the governance of data, and an efficient process that allows us to do this, under scrutiny, which is in the public good.

**Q528 Chairman:** Are you suggesting that we need to re-look at the whole area of the current legislation relating to information about individuals and the governance relating to fulfilling their registration? Are you asking for a re-visit of the whole area?

**Professor Morris:** I am led to believe that it is more the inconsistent interpretation of the current legislation. Speaking to the Information Commissioner, there is latitude within the Data Protection Act which would allow us to come up with a policy which would allow this efficiency. Rory has my recommendation but my other recommendation is that I think we have a duty to communicate what we are doing better to the public – so allied to a public consultation, informing them of the benefits and the reasons behind our activities.

**Professor Porteous:** Could I add to that an endorsement of everything that has been said. Also, if I had an ideal vision, it would be that this would be seen as part and parcel of a welfare health state. This should be seen as part of the NHS function. Indeed, anecdotally, the feedback that we consistently get from participants could be paraphrased along the lines of, "But I assumed that is what would happen to my medical records: that they would be fully and properly used," rather than that there was this very complex process of getting that information. There is an expectation amongst the general population that what we are doing

through biobanking is part and parcel of normal practice. I think that is what we should be aiming for. It should be something that is seen as part and parcel of UK PLC and UK NHS – whether it is Scotland, Wales, England, or Ireland. That is my hope for the future. Anything that achieves that is good in my book.

**Chairman:** On that note, could I thank you all for taking the trouble to come and help us with our inquiry. If there are any issues which come up that you think might help, please feel free to write in. Whatever you say will be used as evidence and therefore will be recorded as part of our evidence. Thank you.