Thanks very much, Sarah, and I'd like to add that. Thanks to the participants but also I want to thank all of the people, the researchers around the UK and internationally, who are continuing to help to make UK Biobank even better for research. For those of you who are not familiar with the background, as Sarah said, half a million men and women aged 40 to 69 were recruited from around 22 areas in Scotland, England and Wales between 2006 and 2010. They gave general consent for all types of health research on the understanding that there would be no feedback of individual results to participants. But we spent a lot of time informing participants about the ways in which you and many others are using the resource to help improve health around the world.

We have a lot of detailed questionnaire information and before UK Biobank started, lots of groups with interest in particular areas came up with the questions that they wanted asked, came up with the measurements that they wanted made, and we tried to include as many of those as we could. We've been looking at ways to extend that information over the past few years. We also had biological samples stored, as you know, for future assays. At the moment, what we're really focussing on and there'll be some discussion about where we are on that. It's around the follow-up of health outcomes through linkage to health records as well as with direct contact with participants. So about a third of a million of the participants we have email addresses, which again facilitates contact with them in order to ask them more questions about themselves but also about their health, trying to extend the range of information we have beyond what we can get from passive linkage to health record systems.

It's available for academic and commercial researchers worldwide and that was an element of the consent without any preferential access and with a limited exclusive access for data that it generated by research. We're increasingly seeing people coming along and essentially becoming funders of UK Biobank by turning samples or imaging data into accessible derived variables so it can be used by other researchers. It didn't start so well. There was a lot of anxiety at the beginning back in the early part of the century, around whether this was a good idea. It's true, it took a while to get going. The UK Biobank was established in 2003 and by 2006 not a single person had been recruited, but they were soon rapidly recruited. Then we did repeat assessments so that one could take account of variability over time, linked into the hospital data, started imaging of 100,000 people. Paul Matthews will talk about our future plans there as well. Then increasingly, more and more data became available and more and more people used the resource.

The linkage is absolutely critical, obviously, in a prospective cohort. But it also tells us about the health of the individuals before they joined, so particularly as we are able to link to cancer data, to hospital data and increasingly to primary care data, we'll get richer and richer data about people's antecedent health as well as their health in the years to come. Although there may be a view that the UK has a single health service, it is devolved and the record systems are devolved. Even within a particular country they are accessible through different sources. So it is non-trivial to get the linkage data, but there is enormously rich data there. We have to deal with a system in Scotland where there are about 36,000 people, in Wales, 21,000, and in England where the bulk of the participants are.
We get regularly-updated information around death and so far, there are something over 20,000 deaths for which we have death certification information. For cancers increasingly rich data about the cancers in terms of stage and grade. As we see for example sequencing of cancers rolling out within the NHS, we are hopeful that we'll be able to link into that. We're also looking at the possibilities of actually accessing the tumour samples so that they could be analysed in the future. So far there are over 120,000 either prevalent or incident site-specific cancer cases. The hospital discharges hundreds of thousands of different kinds of discharges, thousands of different disease cases, all ICD codes.

Then primary care, which is really the jewel if we can get it, we have primary care linkage for about half the participants, pretty much all in Scotland and Wales, and one of the large sources of data in England. We're at the moment working very hard and Mark Effingham will talk about how we've been helped by NHS Digital, by the Royal College of General Practitioners, and the BMA and others, in working to access those data in collaboration with the primary care system suppliers, EMS and TPP, who again have been extremely supportive. The primary care data will give us even richer information about prescriptions, diagnoses, laboratory tests that will allow us to become even more specific about health outcomes.

Looking at deeper phenotyping, we've remote monitored 100,000 with accelerometry data. We're now imaging participants in three centres, with a fourth centre to be opened shortly, so Newcastle, Reading, Manchester and soon Bristol, with magnetic resonance of the brain, with a 3 Tesla magnet, one-and-a-half Tesla of heart and body, DEXA of bone and joint ultrasound, and a repeat of the whole of the baseline so that one can look at change over time in 100,000 of the participants. These data also, like samples, need to be turned into accessible information and there have been a number of groups, both commercial and academic, who are essentially assaying the imaging data and creating useful derived variables around, say, body composition or various measures of the brain.

Paul will talk about how we see this going in the future, and the possibility for additional imaging within these participants. We try to collect the widest range and largest amount of sample. We have 55 millilitres of blood that were collected at baseline and at the repeats in many different kinds of tubes, to do many different kinds of assays. We have urine, we have saliva, we would like to collect stool and we are working on the possibility of doing that in conjunction with the imaging study, subject to being able to get the funding. The strategy for using the samples - and Martin Bobrow may mention this as the chair of the Access subcommittee - is really not to use the samples for a case control collection here, a case control collection there, because although that may be in the short term cost-effective for a particular researcher, as a resource what we want to do is generate data across the whole cohort that can then be used by many different researchers for many different questions and to ensure that the assay methodology used on all the samples is the same across the whole cohort.

The strategy has been to go for cohort-wide assays. They tend to be very cost-effective when we went out to tender for genotyping of half a million people and said, 'This is how much money we've got.' It was remarkable that we got exactly the same quote from everybody. We can do the lot for that amount of money,
which was a lot less than the price at the time. So when you go out to get assays on half a million people, you do get quite good deals. It minimises depletion, you can improve the quality control and as I said, it supports lots of different comparisons using data that are based on the same assay, done at the same time, with the same quality control.

They're costly in the short term but in the long term, of course they are useful for so many different analyses that they are a very cost-effective strategy for a resource. We have been fantastically fortunate with the Wellcome Trust, the Medical Research Council, the National Health Service and the major charities like BHF and CRUK in the UK, Diabetes UK, who have helped to support those assays. We did haematological assays with the fresh samples as they came in, but we waited until we'd completed the cohort before analysing about 40 biomarkers on all of the samples, and those data were released a couple of months ago, after a very detailed quality control check on them, done in our laboratories in Stockport.

If one looks at the way in which the resource is being used, a lot of the initial publications were around baseline data, so looking at cross-sectional information. Obviously as it becomes prospective, it will change but you can see at the heart of all of this, there really has been a big focus on genetics. I think as we get more and more disease outcomes, we may well see a change in this. But I suspect genetics will remain at the heart of this cohort and the research done on it. We released the data in 2012 and this gives you a sense of the increasing number of research projects over time, with already about 19 per cent of all the research projects done in the first half of this year. So we're seeing increasing rate of use of the resource, but we are concerned that there are parts of the world where it's used a lot and parts of the world where it's not used a lot. If we look at South Africa, Egypt, a few places, in China only 270 research groups - researchers using it. We're really trying to encourage use in China.

Just for those from the States, who might be a bit worried about the 445 over Iran. I hope President Trump is not watching and going to have it in a sanctions on UK Biobank. I should say that that covers a number of countries and we are very careful about who does get access to these data, as Martin Bodborough will comment. Again we're seeing increasing numbers of publications; these are the publications each year and last year, 274 publications. The brilliant thing from my perspective of UK Biobank is, you do all the work. You publish all the data. We just sit back and let you get on with it. It's a brilliant model! It's getting into really top journals with again Nature being one that's been very supportive.

There was a very nice series in Nature talking about UK Biobank and it was very gratifying, and particularly I think for us and for the funders with the early concerns about UK Biobank to see this switch around. A lot of publications on a variety of things, and just a few examples are going to be presented later in the day. A lot of press coverage. Some really nice comments in Tweets that everybody appreciates there is a relatively small team, but in the imaging centres, in the assessment centres, it really does… These comments do matter as they grind through the 40,000th image participant last week, heading towards 100,000. The comments from researchers that the value of the data is really very welcome and encouraging.

Just coming back to the Nature article, I think that this is an opportunity not only to thank the
participants, but to thank the funders again because it was the vision of the funders to build this resource, to believe in genetic epidemiology, to put this amount of resource, funding into and keep on funding UK Biobank and making those data available freely that was incredibly visionary. It was great that *Nature* recognised what they had done in making these data available. I hope that I've given an update on where we are and during the rest of the day, we'll talk about where UK Biobank is going and particularly how it can be helped by you.

The other thing we would like you to think about during the day is; how do we make UK Biobank better? We know that it's not perfect, we know that there are things we could do better. In particular, Carolyn now is going to talk about how we're going to try to make the access process better, faster, simpler and also ways in which we're going to try and make it easier for researchers, particularly in low/middle-income countries, to use this resource. Carolyn, over to you. Thank you very much.

[END OF TRANSCRIPT]