F: My name is Robin Flaig and I'm the UK Biobank data linkage programme manager and I'm going to be discussing how do you follow up half a million across England, Scotland and Wales and hopefully keep you awake in the after lunch slot. So let's see if I've got this right. Okay, yes, this one.

M: It's this thing here.

F: Oh, it's that one, thank you. Thank you very much. So our aim is to identify a wide range of incident diseases occurring in our Biobank participants after recruitment. There are generally considered two methods to follow up Biobank participants and cohort studies generally and they are being described from the participant point of view. So the first is active methods. This is traditionally face-to-face interviews, sometimes postal or web questionnaires and this data can often be processed quite quickly and can be quite costly and time consuming and also make it difficult to get complete follow up because you often are only engaging with those people who are most capable of being followed up. The other is determined passive and I have to say I find this term a misnomer speaking as the person who has to obtain the data linkages. It is not a passive method at all from UK Biobank's perspective.

So, why is passive follow-up even possible? Well first of all, it's possible because all of our participants were registered with an NHS GP at the time of recruitment, which means that we can follow them up through linkage to NHS records and that means that we don't have to interact with our participants. We can just get on with the work. This should be effective and cost efficient and it often is, and I have to say, having done this job since 2012, it is getting better. It is getting easier and the data suppliers, NHS Digital, ISD in Scotland, the University of Swansea in Wales, are all really, really helpful to me and have made this easier and easier. We also obtained broad consent at recruitment and that means that we can link and we do have an advantage over some other projects that try and obtain this linked data. However, there are some caveats about this passive method of follow up through linkage and one is that we can only access conditions that are diagnosed in a healthcare setting and the data has to be good. Those are the two main issues.

So we do both kinds of follow-up. We do web questionnaires. At the time of recruitment we obtained email addresses for about 350,000 of our participants and since 2011, we've been doing online questionnaires contacting these participants and over the years we've gotten better at this and we get higher and higher response rates. I would say this slide is slightly misleading because it has a year of these assessments, but in fact, for all of the ones that have happened, we actually leave these surveys, these questionnaires open with the exception of occupational health and the 24 diet recall. So if a participant wants to see what UK Biobank has been doing and they take a wander around the website, they can fill out these questionnaires at any time. Currently we're doing the dietary preference questionnaires and are actively contacting participants to invite them to fill this out and we have some follow-up questionnaires that are going to be occurring over the next few years including a cognitive function repeat and a mental health repeat, but also looking at new things like sleep and autistic traits.
What you'll notice about all of these topics is, these are areas that it's very difficult to get information on via health linkages, which is why the act of follow-up is also very important and this whole programme of work is run by Naomi Allen and Jo Holliday out of University of Oxford team. So this slide should be a little familiar. I've already showed it this morning. So we receive regularly updated information on NHS datasets in England, Scotland and Wales and you can see that for the data that we release is around deaths, cancers and admissions to hospital. These provide many incident cases of disease in lots of different areas. As Rory said it also provides prevalent cases of disease because these datasets go back quite a long way. One of the kinds of data we receive… Oh sorry, primary care data obviously could add hugely to UK Biobank but I'm not going to discuss it, because that is the subject of the next talk.

So looking at cancer, which in the UK because of the cancer registries the data is really robust and the cases are really well defined. The data is good. We have all this data available and you can see at recruitment what the prevalent rate of these cancers is and the incidents, and what we predict will be the incidents by 2022, but as good as the cancer data is, there are potential linkages that could make it even better. These are areas that we're looking to link to, but it's more challenging because essentially these are regional datasets and so we're working with different teams, particularly in Leeds and Newcastle to try and link to histopathology reports and digitised histopathology slides and maybe tumour specimens.

So what data is available via showcase and here I want to caveat because it's very challenging. These datasets, lots of researchers use them but almost no researchers link them together across the three nations and they are different in each country. So, one of the challenges Biobank had when we received all of this data to begin with was, how do we harmonise these datasets? So right now we have deaths, cancer and hospital admission data available to researchers and we do receive annual updates of all of these datasets. We hope to provide annual updates to showcase of these datasets as well, but there are many other datasets in the UK that many of you are familiar with, some that are more unusual. We are currently in the process of trying to determine which of these datasets make sense for us to process and release, because it was no small thing to process these datasets.

So for example, we've had the outpatient data for quite some time, but because it only holds the kind of appointment that was held and the date of the appointment and very few clinical codes, we haven't had the energy and we don't know if it's worth processing. There is accident and emergency data that we have almost across the three nations and there are many other registries and audits and emerging datasets across the country and these datasets - there are many more that I could list, but trying to keep it to one slide, that are emerging and that people are working to create, such as NHS laboratory data. UK Biobank is working with all the groups who are trying to release these new datasets, create these new datasets and as they become available, as we assess them and see if they add breadth of data or depth of data, we can then plan to release them. So, if you would like to help or make suggestions about data linkage please contact me and if you have suggestions or ideas for web questionnaires, please contact Jo Holliday. Thank you.
[Applause]

[END OF TRANSCRIPT]