

How GP data helps UK Biobank improve human health: Participant Event Wednesday 18th August 2021

Speakers:

- **Professor Sir Rory Collins (RC)**
- **Professor Helen Stokes-Lampard (HSL)**
- **Professor Liam Smeeth (LS)**
- **Professor Naomi Allen (NA)**

Full Transcript:

(RC) Hello everybody, thank you very much for joining us today. The team for UK bank is very grateful for everybody giving up their time to join us and to also to thank our two external speakers, Professor Liam Smeeth and Professor Helen Stokes-Lampard, who are going to talk about the value of GP data. Liam is a GP and researcher himself, and the approach and the ways in which the Royal College General Practitioners and the Academy Medical Royal colleges are trying to help ensure that data are available but only available in the appropriate way to help both research and better health care. Before we start, to the 1700 people have already joined.

Is there a problem? Uhm, I hope everybody can hear me. I just wanted to give you a few tips on how to use the system. I'm just seeing that there's a problem with Can somebody confirm that you can hear me? OK, fine, thank you. So again, thank you very much to everybody who's joined, and so far 1,700 people have joined and we are anticipating more will join during. Just a few tips before we go into the first talk from Professor Liam Smeeth on the value of GP data from the perspective of a GP himself and a medical researcher. At the bottom of your screen there are a number of application tools, little icons. Now all of the engagement tools are re-sizable and movable. So the boxes you can see you can move them around. You can make them bigger by clicking on the arrows on top right hand corner or making them disappear down at the bottom. The little icons will allow you to open the ones that you've closed, or close the ones that you've opened.

If you have any questions during this next hour, you can submit them by going to the icon that's got Q&A, question and answer, and just type in your answers. We've got a members of the Biobank team who will try to answer those questions that they can directly into that system, but there will be others that we will bring out either during the talks or in the discussion at the end. What we're going to do is have three talks of 10 minutes and then plenty of time for questions. And if we need to run over a little bit over the hour because there are more questions than we may well do that. I should also say that all of the slides, and indeed the video conferencing you're joining will be available to watch after the event come a day or so after it will be up on the website of UK Biobank so that you can see it. We are aware that sometimes there are problems with sound, sound quality, so there is also a little icon and if you just hover over the icons you can see what each of them does that allows you to get the what's being said typed out so that you can see exactly what's being said as well

as hear it, if your audio is good. If you have some problems with the system if it freezes or anything goes wrong, then the best thing to do is to go up to the top left of your computer screen and there's at least on my screen about 3 in at the very top there's a circle with an arrow. Uhm, which you can use to refresh and that can often solve the problems, but if you have any problems, put them into the Q&A as well and we'll try to help help you sort them out. So with no further ado, let me hand over two to Liam. And I should perhaps have shown this housekeeping slide. But let me hand over to Liam, who is now going to talk to the topic of GP data and how it can be your value for researchers. So Liam over to you.

(LS) Thank you Rory. I hope everyone can hear me. I'm in a NHS General Practice clinic actually, actually in the middle of clinic it so happens, but we'll give it a go. So far so good. So I am a GP, I'm a Professor at The London School of Hygiene and Tropical Medicine and a Director And I'm going to talk about some of the ways we used data for general practice and its value. Lots of work being presented here thanks many people and many funders.

Now I think what happened many, many decades ago, few decades ago when we started to computerized health records and this happened actually very early in the UK compared to the rest of the world particularly in general practice. And I think it soon became apparent for some of us that we could use these computerized health data for research purposes and hopes to do efficient research kind of quicker and cheaper. But that's not the main thing. The main thing is to be able to do research, that otherwise we couldn't do. And be able to do better research. It's not just about cheaper or quicker, it's about being better. I think NHS general practice data in particular is of great value because of its population coverage. You know that there will be basically GP, Its completeness. It's longitudinal overtime, that means and then it can be linked and it can be linked, for example, for studies such as UK Biobank which massively enriches both the general practice data and the Biobank data by bringing them together.

Uhm, I was actually a postgraduate student studying epidemiology research when there was a fuss made as it raised very big concerns about measles vaccine particularly measles, mumps, rubella vaccine and autism. And in 1998 there was a paper in the Lancet suggesting there might be a link and we saw vaccine coverage fall internationally. Both in the UK but elsewhere as well, and measles outbreaks actually started occurring so bad. This is just data from England. Showing measles vaccine coverage. By the time the 2nd birthday where everyone should have had measles vaccine and MMR vaccine. And you're looking for the magic 90% if you get 90% population coverage of measles vaccine, we know it controls the circulation very well. Once it drops below 90, we're in trouble we're going to start getting circulating virus and outbreaks. And that's what happened. And you can see when the study raised concern in '98 there was a sharp drop off in vaccine uptake and indeed this did lead to outbreaks quite bad. We got idea that we could use general practice data to answer this question, address this question. Can measles vaccine and MMR vaccine cause autism? And see similar studies done in the USA and

Denmark and these studies were really only possible because of electronic health records. And because of computerized data from general practice and if you know we haven't had that data that we've been going out and collecting all this data, for example, we still be there. Studying this question, we wouldn't have finished study by now. So we did a study and showed this was really no effect at all.

Going on to that study is this American explains this graph simply to show that what you're looking for is a kind of harmful effect. If all the diamonds and squares were over here. And as you can see, they're all hovering around as a tiny, but not trivial possible protective effect. It doesn't really make much sense, but the important thing being there is no harmful effect demonstrated across all the different studies including our own. And when we, this slide it's one of those slides it looks like you know, I published my study and saved the world. It wasn't just me, it was a global effort going on. What mattered was that these international studies, based on general practice data, really did show there was no link at all with autism. And there was this massive effort to restore confidence in the vaccine, and we saw vaccine coverage climb back above 90% saving many lives.

This is another example. Of showing the value of general practice data. This is some work we did looking at heart disease in relation to how well people with diabetes is controlled. So HbA1c is a measure, a long term measure of blood sugar and basically if you had diabetes in the lower end is better and higher end worse, more out of control. And these are the risks of heart disease that people had and their risks went up enough, depending on how bad their blood sugar control was, if the blood sugar control is quite poor you got a lot more heart disease. And that we already knew. One way we established very clearly but we did already think that. What was interesting about this study was because we had general practice data we could look at this other group who have very, very high risk of heart disease. By far the highest. And what was their blood sugar control? And we could only study this because of general practice data and their blood sugar control was missing. They haven't been to the doctor and haven't had their blood sugar measured and assessed. And the only way we get at this group, who turned out to be the most important group on whom we should be focusing all our efforts their data controlled were the people who weren't turning up at all to get it measured and we could only do this study because of general practice data.

Just turning to COVID briefly at the beginning of the pandemic, we thought about trying to use general practice data to help with the COVID efforts and we had big collaborations between my own institution and the University of Oxford and lots of people helping, particularly software providers, TPP and EMIS health into curating, look after a general practice clinical records and mostly general practice records in the UK. And we did this work on behalf of NHS England. And we basically linked up. As I said, one of the joys of general practice data isn't simply that it can be linked to other sources of data, it can be linked to a general practice data, COVID testing data, hospital admissions, intensive care admissions, death data and all sorts. Initially on 24 million now sent it to 55 million. And what this meant was that very, very early in the pandemic, certainly way earlier than we ever could have achieved without this

amazing ability to use general practice data, we were able to establish which factors were associated with dying from COVID in the study of 17 million adults and we somehow or other managed to deliver in six weeks. Uh, published in Nature, it really guided therapy around the world, and you can see this graph which is to show you how we presented the results, and one of the funny things that happened within about 10 seconds of publishing this result, somebody clever turned into a Star Wars graphic, you can see the death star, and these are the little Star Wars planes which I didn't know, but it was quite nice and was done in about 10 seconds by someone super clever. More importantly, what we showed was, for example, that a recent history of cancer, in particular things like leukemia, and lymphoma, but also that poorly controlled diabetes were very strong risk factors for dying from COVID and this helped guide preventive efforts and policy on shielding and subsequently, vaccine policy. Uh, we showed these very, very marked age effects, where your risk of dying from COVID went up very, very markedly with anything again is directly impacting policy. And we showed very gripping, sad stories around the increased risk seen in non-white groups around the UK and so that this gradient where people living in poorer areas with much higher risk of dying from COVID.

And I was saying all this work was only possible, because we could access general practice data and because we could use it to study these things. So I think general practice data can help us answer some really difficult questions. I think it can help us do that work better, I think it can help us do it quicker and cheaper. And the key of what we need if we're going to make best use of these data and link to these data studies we clearly need to look after the data and confidentiality and security are absolutely at the forefront, I think of researchers' minds. No one wants to see their own or anybody else's health records missing, published or leaked anywhere. And I think we have to maintain public trust that the data being used for good purposes that will benefit humanity, are being well looked after and treated with respect they deserve. And I really do think that as a community we can hold these principles here and use these data and get the best out of them while being very careful, and really appreciating that these are highly private data and really looking after them. That's it from me.

(RC) Thanks very much, Liam. I think some people are having a little bit of trouble with the sound. I hope that you can hear me clearly. If you are having trouble with the sound then one of the things to do would be to use the captioning system that will help by showing the words are being said as well. So I'm going to hand over now to Naomi Allen, who's UK Biobank's Chief Scientist and she is going to talk about how GP data is so important for UK Biobank and the context for that is the inability of us to obtain access to GP data for UK Biobank participants over the last 15 years, despite your consent and she'll perhaps give some examples of how valuable those data can be and indeed have been during the period of COVID-19. So Naomi over to you.

(NA) Many thanks, Rory. So over 10 years ago now, all of you provided consent when you joined the UK Biobank project for us to link to your health related records so that we can follow up your health overtime to find out who develops certain

diseases and who doesn't, in order for researchers to conduct health related research that's in the public interest and this research is being done by scientists all over the world. So at the moment, we are currently linked to national death and cancer registries in England, Wales and Scotland and also national inpatient hospital admissions and critical care data from all the three devolved nations, and over the last year or so we've also been linking the COVID-19 PCR test results to facilitate COVID research, but we still do not yet have any comprehensive access to GP data.

So despite the fact that all of you are registered with the GP at recruitment and all of you have given us consent, explicit written consent to link to health records, why do we still not have access to GP data more than ten years on? Well, there are several reasons for this. So unlike other health record data such as death, cancer and hospitalization data, up until now there's been no single source of data access. The data are controlled by individual GPs and those individual GP practices have had to provide approval to release your participant data to UK Biobank for the purposes of research. Now as we all know, GPs are very busy. They've perhaps been unwilling, but most likely most of them have been unsure about the rules around data sharing. And whilst we've made it as easy as possible for GPs to provide approval to release the data to UK Biobank for research purposes and even provided financial incentives for them to do so our pilot work has shown that only about 20% of GPs would give such approval and as such for us this is simply not a viable approach to obtain GP data from all of the practices from around the UK for all participants.

So this is where NHS digital comes in. So NHS digital are a national custodian for health and care data in England. It's an organization that collects and shares health, administrative data for research and also for health care planning purposes, and has done so for a very long time. So UK Biobank, we receive data on hospitalizations, cancer and death records for participants living in England directly from NHS Digital and have done so since the start of the study more than 10 years ago. And in May NHS digital announced the launch of a new and improved data extract to obtain GP data for the same research purposes, and it is worth reiterating that NHS digital is a longstanding organization, they have well established infrastructure to deal with the issues around de-identification. There are rigorous approval processes and a robust regulatory framework through which us and other research organizations have to apply to receive these data for specific research purposes. So when the data announcement was made about this data extraction a few months ago, there was a small but highly vocal minority of campaigners who were frankly peddling quite a bit of misinformation about the dangers of the NHS will sell data to companies for profit, including insurance companies, and there will be a mass infringement of privacy by this so called "NHS data grab", and were encouraging the public to opt out of this data extraction program. So in response to this, representatives from UK Biobank and also other members of the medical and social research community held a press briefing to explain the wider benefits of GP data to accelerate really important research.

In essence, the access to patient data can save lives and that such misinformation could really scupper vital research if sufficiently large numbers of people decide to

opt out of the data extraction. So what is it we're actually talking about? What is included in this GP data set? So the first thing to say is that it's a single standardized data set. It covers all patients in England, and so for us it will cover all UK Biobank participants that live in England and is coded in a single format, which means it's much easier for researchers to analyze it more efficiently and quickly. Some data goes back to the mid-1990s, but for most patients the data will go back to the mid-2000s, but for this data extract, some items will only go back for the last ten years or so. It only includes coded data, so this will include codes related to signs and symptoms you've had any clinical diagnosis, lifestyle factors, any vaccinations you may have had, referrals, and of course prescriptions. It contains no free text or attachments. For example, letters from consultants and so on. So just why are GP data so valuable for cohort studies like UK Biobank and other studies as well? So access to GP data will increase the numbers of health outcomes that can be detected because there are many events included in the GP data that do not actually lead to hospital admission, let alone death and will also increase the range of health outcomes that can be detected. Because there are very many conditions that are typically only managed in primary care. If you think about conditions like arthritis, depression and anxiety, diabetes, hypertension. Most of these are typically only managed by your GP, and so if we don't have access to the GP data for research, then it goes to show that research into these important conditions is very much limited. So this table just shows the number of cases identified in the UK Biobank cohort by the end of last year of a handful of conditions identified through hospital and death data, and then those numbers in red in the second column identified when you would also include GP data. So when you look at something like a heart attack where most patients will end up in hospital, you can see that the GP data doesn't actually add a significantly more number of cases, only about 6% of total numbers of heart attacks are only found in the GP data, but if you look at conditions like diabetes, chronic obstructive pulmonary disease or Alzheimer's disease, you can see that between about 40-50% of all cases of those particular conditions are only found in the GP data. And then when you go to the other end of the spectrum, when you look at something like depression, anxiety, rheumatoid arthritis, about 80 or 90% of all of those cases are only identified in GP data.

So if we don't have access to these GP data, then research into these conditions that cause a lot of illness, debility and disability is highly limited. So GP data also provides information about underlying medical conditions, so it will enable researchers to look at how important other conditions that you have are on the outcome of a particular disease. It also provides information about prescriptions and investigations you've had, which will allow researchers to look at how important various treatments are on progression of a certain disease. So at the start of the COVID pandemic, the government realized very early on just how important GP data are for health care planning purposes and for research into this new disease, and emergency legislation was introduced that enabled UK Biobank to access GP data directly from the GP software system suppliers and this GP data enabled and continues to enable vital research into the clinical determinants of severe COVID-19. So, for example, research using the UK Biobank data has shown that conditions like obesity, prior kidney failure and previous infections all lead to a significantly

increased risk of being hospitalized for COVID-19, research which just simply would not be possible if we have not had access to GP data. But perhaps most importantly, what this work is highlighted is actually what's been lost in us not having GP data to enable research into all those other conditions, not just COVID, but arthritis, depression, dementia, asthma, eczema, migraines, hypertension, diabetes. I mean the list goes on and on. Researchers cannot do robust research into the genetic and lifestyle determinants of these really important conditions without having access to GP data.

So finally what happens now? So as you probably are all aware, the data extraction program that was announced back in May has now been paused to enable a wider consultation between yourselves public and the GPS to make sure that the public's properly informed about what's going to happen, what the data consists of to communicate the safeguards over privacy concerns, and to enable you to make an informed choice about the use of the GP data for research purposes. So I'll leave it there, and of course very happy to answer questions at the end.

(RC) I think I should step straight in then. Is that right? Thank you very much. Yeah, thank you very much for that Naomi. I hope that the sound quality for Naomi's talk was good. As Liam said he was joining us from within the NHS system in his clinic. So it was a bit more difficult but yes now thank you again to Professor Helen Stokes-Lampard for agreeing to talk about this from the perspective of the Academy of Medical Royal Colleges and of course she was also the previous chair of the Royal College General Practitioners who have been discussing for a long time how GP data could be of value. So Helen over to you, thank you very much.

(HSL) Thank you very much, Rory and Naomi and all the team at Biobank. It is lovely to be with you all, but you know despite how long this pandemic has been going on for, I still find it very peculiar being doing this sort of thing remotely. I very much relish face to face interaction, although I don't think we'd have as many of you in a meeting room somewhere.

So good afternoon everyone. I'm Helen Stokes-Lampard. Now I'm actually also a GP by profession. I'm based in Litchfield in Staffordshire. I'm also an academic. I'm technically a professor of GP education at the University of Birmingham. Although I am seconded away from that at the moment. I also wear another hat and I'm chair of the body called the National Academy for Social Prescribing, and that's a new charity that seeks to explain what social prescribing is making a noise about a movement that's supported by the NHS in England, which is the first major health care system in the world to do so, and ultimately ensuring that everyone can access the support they need to help them live their best life. Whatever your medical needs are, and so our vision there is a social revolution in health care. But of course Rory asked me to speak to you because of my main role that I'm currently the chair of the Academy of Medical Royal Colleges, which is a body you may not be familiar with, but essentially what we do is bring together every different medical discipline. All the Medical Royal Colleges and medical faculties and we achieve consensus on

standards of care for patients, standards of education and career progression for doctors. And we also speak out on behalf of doctors on issues that affect us all.

Now in this role I'm privileged to meet regularly with leaders from across health and care, and I have frequent briefings and what's happening and why, and so in turn I can influence these things and it is my greatest privilege to bring together all the disciplines large and small, to ensure that the challenges one area of medicine faces are supported by all the others and patient data and how we best handle it is one of those challenges and responsibilities. Now as doctors we really understand the potential sensitivity of medical records and the trust that our patients place in us to keep information confidential. And having told you a lot about what I am, I also want to be clear about what I am not. I'm not a spokesperson for the government. I'm not a spokesperson for NHS Digital or NHS X, and I'm not here to defend anyone. I'm just going to try and share what I know and my perspectives. But I think I have to refer to the pandemic first. The current climate in which we operate is what I call a COVID tinged world.

Now it's a world where inequity and the social determinants of health have a very tangible and immediate impact on survival from infection. When I took over as chair of the Academy in July last year, we were at the end of the first wave of the pandemic and at that time the UK had just under 300,000 recorded cases and 41,000 people were dead. Today we're still in the turmoil of a very painful third wave, despite the fact the newspaper seemed to forgotten this. We have over 6 million confirmed cases in the UK and over 154 thousand people have COVID on their death certificate. Yesterday, we added another 170 deaths to that total. To put that in perspective that is 3 coach loads of people died yesterday from COVID. It's no headline anymore. Worldwide there are over 200 million formally diagnosed cases with over 4 million people killed by the virus. And those are the ones that we are sure about. But on a brighter note, last summer a vaccine seemed an eternity away and now almost 90% of the UK eligible population have had one dose and 77% two doses and worldwide we're heading towards 5 billion doses of vaccine. And by the way, please, if you know anyone who is unvaccinated, who's eligible particularly pregnant women, please urge them, beg them to take up the opportunity. So we are really in crazy unprecedented days. But as you've heard from the others, it's the UK's superb medical research that has led to a transformation in the way we treat this new disease. And enabling much of that research is the power held within high quality medical data.

Liam, Naomi, Rory have all told you this. I won't repeat it. But it's been mentioned that I used to be Chair of the Royal College of General Practitioners and several years ago we discussed very positive plans to improve the way that confidential patient data could be responsibly used by the NHS to improve data security and safety whilst boosting research opportunities and NHS planning, and a great new plan was developed to make this a reality. And a few months ago I was alerted to emerging national concerns about these much anticipated improvements, and by now this was called the general practice data for planning and research program. Like it's a clunky name GDPR. I didn't come up with that. But the problem was, as Naomi

highlighted, there was speculation in the media and on social media targeting the public and doctors. That was almost entirely negative and almost entirely wrong. But it did raise some uncomfortable and important questions about the project that definitely needed addressing. There's no doubt amongst health care professionals that access to robust and extensive anonymous health data is essential if we're going to tackle the health challenges faced by the population. But if the proposed solution was losing the trust of GPs, it certainly wasn't going to retain the trust of the public. And in my role in the Academy, I was asked to help find a way through. So I canvassed a lot of professional opinions and I made recommendations to the former, now former, Secretary of State for Health and Social Care, Matt Hancock and his team. And after a lot of lobbying and requests from many quarters to slow down the process to ensure greater clarity to make some fundamental improvements and build confidence, a few weeks ago the government gave very clear assurances that no data will start to be collected via this new system until four conditions have been met, and Naomi alluded to them, but I'll give you the detail of them just to be specific 'cause I know some of you are here for this detail. And these four conditions are first that if anyone has chosen to opt out of sharing their data that their choice has been recorded and implemented so there's no backlog of opt-outs. Secondly, if somebody chooses to opt out of data sharing, even if their data has already been uploaded to the system, it will be deleted. There's no deadline for opting out, in the same way there will be no deadline for opting in again, and certainly we would hope that people will change their minds and opt back in once they are confident that the process is robust, the data usage is legitimate. I've seen questions where people are saying who's going to sell my data. I hope I will reassure you in the next minute or two. There is no sale of data going on. But third, that patients must be made far more aware of this scheme through a brand new campaign of engagement and communication. So please look out for it. It hasn't started yet. I suspect it'll be a few months before it does, because it's going to take time to do this well. But finally, something called a Trusted Research Environment will be developed and implemented by NHS Digital. Now Trusted Research Environments are highly secure systems intended to hold large, secure data sets to give people confidence that their confidentiality will be maintained, and it's what the office for National Statistics have done for 16 years now. And it's the gold standard internationally. So this means that NHS patients will also have this highest level of assurance. But researchers will be reassured to know that this new system will be designed to make their lives easier too, as they will be involved in developing it. And Joe Churchill, one of the health ministers, has stated categorically that NHS patient data will never be for sale for commercial organisations to use or abuse. Lord Bethell, another health minister, has also clarified that the GP data being referred to is coded data that was mentioned earlier. It's not the free text notes that GPs like me often used to add richness and personalization to records. So it's coded data. It won't be sold and it will be used in a very responsible way and in an equitable way.

So I hope this is reassuring to those of you who are here to learn about those issues and I'm happy to expand in the questions. Now of course you are all here as research participants who have already given your informed consent to data sharing. So you've demonstrated your altruistic streak and you've also demonstrated that you

understand and value the importance of medical research and just to be clear, as has been said already, if you have actively consented to data sharing for research, this is separate from the GDPR challenge. So thank you for what you have done and continue to do. Of course, the technical issues of accessing the data that Rory and Naomi are concerned about are a parallel issue. But you've given your consent and that is and should remain separate to GDPR, so I'll wrap up in a minute, but I want to finish by saying that I believe this pandemic has created an inflection point of the NHS. With a backlog of five and a half million people currently waiting for some sort of investigational treatment in England alone, we cannot continue to work in the ways that we've got used to do. We need to rethink a few key areas. We need to think about how we work in ways that minimize harm to the planet. Our environmental responsibility is something that every single one of us must take seriously. Where care is delivered, what should be done remotely like this? What's best done locally in our communities and what should remain in specialized settings as an example. Who delivers care? We've now got myriad new health care professionals working alongside traditional roles like doctors, nurses and physios, and these new professionals will be more visible and we need to find ways to explain to everyone how they fit into the delivery of high quality care. And the increased personalization of care is vital from high tech end like genomics and robotics that have been mentioned. But to the social and spiritual ends of our lives that impact on our individual health journey and our longevity. Now technology always comes into this sort of conversation. How amazingly well we've done at embracing health data. And we have fancy dashboards and massively improved remote connectivity with patients and new apps and tools at our disposal. And that's absolutely right. But it has also graphically illuminated the digital divide. And it's also made many clinicians like me reflect on what it means to be a good doctor. To consult remotely, especially for complex, nuanced situations is a lot harder. It's slower, it's more difficult, and it's less satisfying for both parties in many occasions. And as clinicians we tend to over investigate over refer to play safe to overcompensate if you like, and that's ultimately wasteful, and it can create unintended harms for patients. So as we look to the future, technology will absolutely play a larger part in our routine working, but we'll better understand its place and its limitations. So thank you all for joining the webinar and for being a vital part of the Biobank program. Each and every one of you is making a real meaningful difference to the future of healthcare research and I look forward to the discussions. Thank you back to you, Rory.

(RC) Thanks very much Helen. And to the previous speakers and I know there were some problems earlier on with the listening to it. I hope that you can hear me clearly and that the captioning is working. And of course you can have a look at the presentations again when they go back on the website. I'd like to throw some questions out to the panel if I could, and one particular one that I'd like you to perhaps discuss because I think it's relevant to a lot of the questions that we've had from the audience, and that is the balance between the risks of making the data available, which there's been a lot of emphasis about anxieties of, how you might be able to identify people from deidentified data, versus the benefit of the data being available, actually both, not just for research, but of course for providing better health care through the NHS itself.

I don't know whether Helen I could ask you to perhaps kick off on that? And maybe Liam and Naomi can kind of talk about those, that the risks that people are concerned about that if you like the sins of Commission of doing something versus the benefits that we're currently not getting by having these data available.

(HSL) Yeah apologies, I had a little bit of interference on the line as you started there, Rory. But I think what you're asking about is, is this sort of reassurance and what persuades people to get involved in this. Is that correct?

(RC) Well, really to talk more generally about at the moment the emphasis is on the risks of access to data. And the risks of people being re identified. Now, As Naomi has said, the death, the hospitalization, the cancer data are already being used very widely. But there's been a big concern raised about the risks associated with the same thing happening with the GP data and just trying to kind of balance that with what's being lost by not being able to use those data for both for research and for better health care.

(HSL) So the risks of re identification are actually very small. The scale of datasets being used is enormous and if you've got large scale datasets, the chances of being able to identify individuals gets a lot less. So if you're talking about, you know the things we've been talking about mostly are common diseases for which there are hundreds of thousands or millions of people experiencing the risks of deidentification get greater when you are dealing with very, very rare things, or you're in very defined geographical areas and there has been examples of how people have used you know so called deidentified data and re anonymous, and re identified them. Actually the standards that are currently applied are very, very high and the levels of assurance are high. But it is because we feel it could be better than the GDPR approach is being taken. And of course with bodies like Biobank what you're doing is holding data incredibly secure, regulated ways, so people, so I'm always conscious we've got two separate things here.

We've got the wider question of GP data and we've got the Biobank standards and the research methodologies and I don't want people to confuse those, and you know people have consented within a study to put something in the wider research thing, the rules are different. The use of a Trusted Research Environment in future will completely remove the possibility of re identification. It will be complete anonymization and nobody will be able to get the key and that is what happens with the ONS data at the moment and the Office of National Statistics pulled together, I think it's over 80 different datasets and can link them. But nobody can get access to identify who the individuals are, and that standard will apply. Researchers are anxious that that's complicated for them at the moment, but building an ease of access for researchers to use that in the most user friendly way will help them. I don't want to get too technical and hope I've answered it enough, Rory, to let the others pick up.

(RC) Thank you. Liam perhaps you could continue that general point.

(LS) I'm trying a new microphone with some scary looking headphones. Yeah, I hope it's better. Yeah, a really key thing to say of course is that re identification for sort of bad reasons or any kind of malice is actually against the law. It's a criminal offence and that's quite important. This is not something that happens casually, and there are many, many safeguards against it, and it's one of the reasons we haven't seen it happening over the years, and I think as Rory says, the benefits the existing benefits are so huge and the thought of turning off the use of this data, even the existing use of this data, it really does illustrate what would be lost, and it helps illustrate what we can gain through by using these data. I mean, just in terms of running the NHS, making sure the care is good and appropriate, and going to people who need it and how much care we need to provide, you just take COVID as an example. Knowing what is going on around the country in terms of infection, looking at the vaccine, who is getting it? Are they the right people? What are the adverse effects? What are the beneficial effects and what impact is it having on the pandemic? These are all things that we've been able to do only because of these kinds of data and the thought of not being able to use those things, the harms would be enormous, so I think it's quite important, as Rory says, to really think through this isn't a kind of yes, as bonuses and additional things we can do with these data, but the thought of not using them would actually cause great harm to human health and the risk of identification are both tiny and that's one of the reasons it hasn't happened to date. But as I say, the safeguards are enormous as well.

(RC) As a GP Liam, and one of the questions that come up from a few people as well, you know, I'm aware some participants have said that some of the data that the GP has on me is not accurate, is that going to invalidate its use in studies like UK Biobank. And it's the reverse if the data are actually available and being used both for research and for your health care the data themselves will get improved. Yeah. What are your thoughts?

(LS) I mean, I think that's right. The data are imperfect, all data are imperfect, but data clearly that are collected for health care, routine health care, some of them are very some of the data are really accurate, some are slightly less accurate, and it's one part of good research using these data to think about issues we could talk about data quality, meaning how kind of accurate and complete it is. And actually one of the things we can do using a study and linking into a study like Biobank is actually it gives us a great opportunity to see how good bits of different bits and pieces of the data are by comparing it, for example, with the data we have in GP records with the data that many of you of course, in fact, virtually all of you contributed for example in the assessments you've done for Biobank where you've actually seen researchers and had measurements taken for example. And by comparing these data sources, it really lets us see where the strengths and possible weaknesses of the data are, and it's important we're not, you know, we're not blind to those possible weaknesses and always have them in mind, and there are going to be weak spots in data. And what we always have to do is try and find them and make sure we're relying on the strengths in the data and be aware of that. So the data are not perfect, but it doesn't mean they're not useful.

(RC) Yeah, one of the questions that also came up that I noticed was that NHS digital was seeking to get access to coded data and was that involving a lot of extra work for their GPs and I think to reassure people that GPs have been coding their data for a long, long time. And in fact GPs I think were ahead of the game in many ways in terms of using coding systems and using electronic ways of handling data. So really, what's being sought is just a subset of the data that the GPs already have, which they have already coded. So in many ways will reduce their work.

(LS) Absolutely, this is no additional work load at all for clinicians. And, as Rory says, increasingly UK general practice lead the world in coding clinical data, so coding means you know you turn up to the GP with acute bronchitis and rather than me just writing down acute bronchitis, it actually gets coded. So there's a code, I can't remember what the actual code is, but it'll be a letter or number or letter underscore, another number or something, and that means acute bronchitis to the researcher who get that data. And we know that the coded data are both more useful for research, they tend to be more consistent, but there's no additional workload whatsoever for the clinicians, no.

(RC) Can I just go back to Helen before I go to Naomi because one of the questions that comes up I suppose that very basic question of trust. Why should people trust the government not just to sell the data? Why should they trust them?

(HSL) You know it's a really good challenge under several questions. I spotted scrolling through about trust and sale of data. We trust the government all the time. I go back to the Office of National Statistics and Trust and surveys and censuses and our tax details and all our financial information is all through the government and this you know, we despite our cynicism and our criticism of the government, we have some of the highest standards of integrity and transparency in the world when it comes to our government and particularly our civil service. And for all the criticism that we leveled at them, and the criticism usually is leveled at individuals who may not be playing the game properly, but as a system it is a very high standard and we as the electorate hold the government to account. We hold our politicians to account and they will not be elected if they messed this up badly. I mean, at the end of the day, that is our ultimate sanction and we can use it. And in the NHS and healthcare we have an enormous power, an enormous voice, and if we speak out loudly and strongly, I'd never underestimate what we can achieve if we choose to really put our force behind something so to have made sweeping and very profound statements as I referenced with Joe Churchill, said we will never sell NHS data. That is a very strong assurance that we should cling to. Use of truly anonymized data to plan our services better is a very, very good thing, and patients, my patients, are frequently horrified that the data isn't used more to plan services better to predict how many doctors and nurses we need, to how many hospitals and GP surgeries we need and so on.

So there is so much more we can do if we use the data better, but those safeguards must be massive and I guess you know the reason Liam and I are here is trying to help reassure people and get the word out that this is a good thing and very well

intended. And we will use everything at our disposal to make a noise if we don't think the right stuff is happening, as we've done in pausing this GPDPR initiative, and there was a great question I spotted scrolling through and please indulge me. Somebody asked Isn't it disingenuous to call it GPDPR when we've got GDPR, and just to be clear GPDPR, the general, ah I'm getting my tongue twisted myself, so there's the national data protection regulations: GDPR, totally separate from the project General Practice Data for Planning and Research purposes. 'cause the first thing I did was get them confused myself. It's an outrageous oversight on somebody's part when they came up with the acronym, which is why I sort of slipped into my whole please don't blame the acronym it would be awful then and I'm thinking I probably a new name for the project will come out in the near future. So by the time we come to the patient engagement campaign will have sidestep this nonsense of getting acronyms mixed up, so thanks for flagging that and thanks for giving me an opportunity Rory I am not a spokesperson for government. I'm not here to defend them but I will say we have got very high standards and having a National Health Service is very helpful in this sort of thing.

(RC) So one question actually that also came up, Helen was around if people have opted out of the GPDPR system, what's happening now, given the pause, do they have to opt back in? What do they have to do? Because of course opting out means that despite consent, UK Biobank would not get their primary care data.

(HSL) Yes, so you can certainly opt back in at any point. So currently it would be going back into reversing the process that you already did. Which is usually back to a GP surgery to fill in a new form to opt back in, and in the near future the idea is to make it far easier to opt in and out properly using the NHS app. There will always have to be a backup system for people who aren't technologically enabled or at digital disadvantage, and so it is intended to be easier and more transparent. One of the big problems is that people get very confused of what the various opt-outs mean anyway. And I think people are frequently horrified when they realize they opted out because they were concerned about the noise in the media, which is completely understandable, but actually they do want their data to be shared for purposes, but instead they opt out of sharing. They put for purposes of helping their health, so your data can't be shared with the hospital trust for the ambulance that's picking you up, but I mean that is really, you know, in this day and age for the ambulance not to know what your allergies are, for example, is fundamentally dangerous.

So we need to make it much clearer for everybody to know what's going on to understand the reasons why so people can make informed decisions, not panicked decisions, and then opt back in a very straightforward way. So please, if you are considering opting back in, of course, I would urge you to do so, and I hope the reassurances of what's happened in the last month or two and what we're saying today will help you with that and everything I'm saying is in the public domain, if you want to know more about this, there was an evidence session to the Health Select Committee, which is now chaired by Jeremy Hunt, the former Secretary of State for Health, where he interrogated people like Lord Bethell and people from NHS Digital, The National Data guardian on this very issue. So if you are so inclined the

Parliament channel the BBC Parliament channel has it all there and a quick search on GP data you will find it, and it's the second half of the session where the most interesting bits are with respect to this, I would suggest.

(RC) We're getting to the end of the hour, and I know Liam may need to leave, but there are quite a few questions so I am going to just keep that going a little bit longer. If other Members can stay on. Just to follow up on that Helen, there's some questions or confusion about the timing, the timetable of what's happening, and still anxiety that the data are going to be downloaded in September. Where is this communications plan? And can you outline from your perspective, what do you think the timeline is likely to be?

(HSL) Sure, so first of September was always a bit of a spurious deadline anyway, because it was only that the process could start from there that's gone completely. Those four criteria that I said have now got to be met and there will be no new hard deadline, and the communications plans being drawn up at the moment. My understanding is that in a couple of months there will be a discussion with public and patient representatives and bodies about what a good comms plan would look like, and I can't see that kicking off, late winter I would have thought would probably be the soonest for something about sophistication and nature, and building a Trusted Research Environment and major technological challenge is going to take many months, so we're talking probably next spring at the earliest, possibly next summer, so in the meantime nothing changes. The current systems will apply, which will allow time for any stated opt ins and opt outs to be processed. So I hope that helps on the time tabling, but fear not. First of September is not a thing anymore that is gone as a thing and it was only ever and I do not think there will be a specific data set in future. There is talk now about starting to roll things out, probably in a piloted way, and a regional way rather than a big bag approach to this 'cause it's so important to get it right. We really don't want anymore debacles and public getting panicked and anxious about something which really shouldn't be something to get panicked and anxious about. And there was comments about public.

I'm just trying to look if there's anything else in the questions that I can quickly see I think I probably covered the main ones of it and in terms of hacking the system so the NHS works with GCHQ and just digital world for GCHQ very closely to ensure that NHS data is of the highest security level as Liam has probably had challenges today using NHS tech to connect with the outside world. We've got very strong firewalls with NHS tech and it makes it very difficult sometimes to communicate externally. Which is actually a good thing in terms of the protection and making it hard to get in. There is only one significant challenge that was how that was the WannaCry attack several years ago, which was a vulnerability in old and outdated software which wasn't having ongoing updated protections, and the NHS has now moved away from that. It was a wake up call I think across the system and I remember it very vividly when that happened, and that did affect a few hundred GP surgeries, although no individual patient data was compromised. But it did cause a hiccup, so that is something we can never be complacent about. But if you are interested then the

NHS digital website and the National Data Security Center is worth a look online. It is very impressive and very interesting, thanks.

(RC) And of course Helen these data already exist and these data are already on centralized systems. The primary care data, as Liam I'm sure can say are on commercial systems, run on behalf of the GPs, EMIS being one TPP another and a few other smaller companies. So what ones talking about is something that would be done centrally by government and with perhaps even more controls, and certainly more oversight than is currently the case with the data as they're held now.

(HSL) Absolutely that is very much intended. The bar will go higher and higher. This isn't about bringing the bar down, it's about making things better and more improved for the future with higher levels of assurance and reassurance, and you mentioned the commercial systems just to outline a few in my own surgery we have prescribing data that goes back to 1984, so any drugs we've prescribed since then if you remain taking for that time so you know it is a phenomenally rich data. I mean and the codes and the diagnosis got added later so they have increased with sophistication. But yeah, but they are they were built originally by GPs. Who are, you know, nerds of their time who designed what gradually became big commercial organisations hospital data spaces and the resources tend to be more commercial from the outset.

(RC) And Liam, I don't know before you go whether you want to kind of comment on this because of course you've used such databases. The CPRD database for example.

(LS) Yeah, and it is true that the because of the way, and it was why the NHS you know, and general practices in particular were computerized at such an early stage I have to say it was partly commercial interest and pressure, but what's interesting is that the clinical data in the NHS have been held by the software providers essentially the companies since the 80s even and certainly very almost universally since the 90s. And it really is their number one job to look after that data and that's why there's never been a single breach on these commercial systems, so people's general practice data when I type into someone's clinical record, it goes off into my practice uses EMIS as it happens, goes off into the EMIS Cloud on their servers, and they look after that data securely, and they've never had a data breach. And I mean, it really would be the end of their company if they ever did. Of course. So not surprisingly, they take it incredibly seriously. So the software providers come, you know, I can honestly say having worked with them now for a long time they really, really know what they're doing in terms of data security and my medical record is really precious and private. I can assure you and I entrust it to them, and without any hesitation, and so I really do think that we know how to look after these data.

(RC) Yeah, and again going back to my original question about risks versus benefits. It's not as if anything completely new is being invented now. These data have been

held by centralized systems for the decades and being used for certain research and for audit purposes for decades. Naomi I wonder whether you might want a kind of come in on this issue of risk versus benefits, and perhaps talk about the security around the UK Biobank data, and indeed the Trusted Research Environment that we have built, which, as Helen said it's a non trivial thing to do and we built it largely for because of the scale of the data. But also the question that's come up a few times, if not several times, is Why didn't we tell the participants earlier about the problems we've been having with getting access to the GP data? I think the short answer is we've been working very closely with the RCGP and with GPs to try to overcome concerns, **but is there anything that participants could do now to help UK Biobank?**

(NA) Yeah thanks so in terms of the data security around access to UK Biobank data at any rate, as Rory's alluded to, we've actually kind of built our own Trusted Research Environment. So for those of you who don't know, this is essentially a cloud based database behind a very strong encrypted firewall where researchers can do their analysis in situ, and so the data are not being downloaded onto their local computers, so this means that all of the UK Biobank data, including the GP data, your hospital data, your genomic data, and all the rest, will be analyzed securely on the cloud. Using strong encryption and will not be downloaded on to local computers and so forth. So you got much more control over the data. There's absolutely no personal identifiers in the data, no names, addresses, Date of births, and NHS numbers and so on and so forth. So it just means that these trusted research environments which are now very popular in the research environment, means it's much more secure, safe, efficient, cost effective to analyze very large data sets on the cloud in a safe way, rather than actually downloading the data that you need to your own Laptop.

(RC) So Naomi, just to be clear I mean up till now the data have been downloaded to researchers to work on their own computers. We built this Trusted Research Environment particularly because of the scale of the sequence data is so big that actually it's impractical to download those data. But I think going to Helen's comment about the development of trusted research environments by NHS Digital, I think we are working with NHS Digital to perhaps pilot the use of a trusted research environment with the GP data. As they start to think about how to build the system because it is, as she said, nontrivial to build these systems, and we've spent the last year building one, so we know that, and I think that working with NHS Digital we can help to show what can be done on these platforms, how to make them very safe, but, and it may well be that allows the GP data to be available to UK Biobank within the trusted research environment, even though the other data are still downloadable.

And Naomi, what about the question of how do participants know whether their GP has made their data available? And what is it that participants might be able to do if they want their GP to make data available? What would you suggest they might do? And indeed, Helen, you might have suggestions on this as well.

(NA) Well, what I've done is I've phoned up my GP practice spoke to the receptionist and made sure that you know that I want my data was not opted out for, that the GP

hadn't done it on my behalf either by mistake or deliberately. And if you're unsure, check with the GP practice as to what the status is of your records and if you have, you know if you have any concerns you know you contact your GP and you let your GP know about the issues that you that you might be having, that's certainly what I've done and they've been responsive to it.

(HSL) Just have to say this, the actual GPs don't know. It's there's usually a managing person in the practices that are involved in this or in our practice it's one of the reception managers who does the actual box ticking and unticking, and I have to say when Rory raised with me this issue that practices are not releasing data despite patient consent, I was a bit horrified because we've never considered not releasing patients data when they're specific in research studies, so I confess Rory that I don't actually know the process for getting around it. But if you are going to contact your GP surgery, can I suggest you do so in the afternoons the phone lines are generally very busy in the morning. You don't want to wait for an hour in a queue of 80 people to get through to talk about something that is a technical matter as opposed to needing an appointment on the same day. So just top tip there for contacting your surgery and it may be something that actually your practice will accept an email query on. Most practices have websites and email queries in this sort of thing are extremely well handled in that group. My top tips.

(RC) Yeah, I'm sure an email would, rather than trying to phone the practice, particularly given how busy practices are now having to deal both with COVID and all of their standard work. I mean the GPs more than perhaps any other part of the health service of probably the most torrid time of all. But I mean, I also am a participant in UK Biobank and I feel Naomi's approach of just dropping a note saying I would like my data shared, because we tried very hard and despite the support of the Royal College General Practitioners, despite the support of the Information Commissioner and the system supplies Unison TPP, actually it was very difficult to get the general practices to approve the release of the data, for consented participants, when we wrote to GPs and asked them just to essentially push a button for the system supplier to make the data available for the consented participants. Only about 20% of them did so. And we've really struggled with trying to get further than that. So I think perhaps this is a good time for participants to engage with their GPs in a way that's positive. But clearly it's also important to understand how busy they all are. I don't know whether there are any other questions or comments that Liam did you want to make any other comments? I know you have to go in a couple of minutes.

(LS) No, just to say thank you to all participants, of course, and how valuable the work is and both Biobank and I really do think it'll be so hugely enriched for really years to come by this this linkage GP data and I think it's a huge value, no one's ever gonna pretend there are no risks attached to doing medical research of any sort, but the risks are tiny compared to the massive benefits we know under the years to come. And yeah, thank you again to participants. And I do have to go with very sorry.

(RC) Yeah, thank you very much Liam. Thank you Helen, and thank you Naomi and thank you to all of the participants in this, around 2000 people who participated in this. Do send us other questions if you like by email to UK Biobank if you have any questions or comments, we do want to work out how to address those. I know I haven't covered all of the questions that were in the Q&A. I know the team were trying to cover as many as possible, but we are incredibly grateful to you and all the other UK Biobank participants for contributing to this resource which is making a huge impact on better understanding of health research. So thank you for your participation today. There are a couple of questions that will come up at the end of the session about this event, please do give us your feedback as well, and we certainly pick up on the feedback about sound quality, so hope we got that one. Thank you very much everyone. Bye. Goodbye everyone.