Tokyo Workshop on Genomic Medi...eutics, and Health 20240411-01

Thu, Apr 11, 2024 9:31AM • 58:38

**SUMMARY KEYWORDS**

participants, data, questionnaire, uk biobank, information, cohort, field, instance, included, study, recruited, visits, number, diagnosis, population, proteomics, baseline, questions, people, center

00:00

thought if we gave a chance for students to learn more about data science this year, as you understand from the previous International Symposium, we invited a large number of UK Biobank people, including the time to show us the full picture of UK Biobank and how to use the data, how to share the data and how to analyze and the workshop is in this, the contiguity of this symposium. And so, the topic of this year, or the program to this year is how we are going to use B 35. For our Genomic Medicine study. So, we got three days and our first day as you see we have different lectures coming from the UK by want. As you see Lucy Daisy Daisy will see the last part of the lectures today is by Lucy and Dizzy sitting over there. And at the end of the day, we have Gemma who is going to talk about the the application of you eg via genetic data and data. And then after that we have again AZ and we have another disc and the second thing, we invited people from different biobanks in Japan, to showcase what they are doing and what is the possibility of data sharing and collaborating. And to decide if that's possible.

02:41

Like this not only by advancing Japan

02:56

in the afternoon we have with going to present the cached version, which is ridiculous cobalt for health study and we have different people including McCann to talk about psychiatric researches and some other people from Britain University of Tokyo SATs national centers and the third day in the morning, we have a sort of big lecture on polygenic risk score modeling by permadeath and in the afternoon we have hazard assessment. So this is a very very long session or three days and I hope that everybody don't get tired in the middle of the course and learn things from what we have here and all practical information. One thing which is very important for everybody coming Saturday, is the main entrance of web on the ground floor is closed. So you need to enter a building from underground and we are going to distribute the the map and the supplemental information will be at the end of the day. That if you come from the main entrance of the building of grandpa living today you will get used to

05:06

this building I was here Saturday just in case so I have a co organizer diversity was to say something

05:33

just a small note to say that this year's workshop is a bit different from previous international profile we have more lectures and events last summer practically. So the third day will be a bit more like a standard kind of format that we're using the workshops where you're getting more hands on. So but hopefully again you're gonna find everything interesting information has some of these more advanced methods in the 13. But I would also encourage you to look up because he organizes workshops every year, and so future year more but you can see, we'll cover when

06:28

you do so we'll just start the session. And I'm just going to say

06:41

yes, so the very first hope you

06:46

will see from the speaker bipod is going to give you the idea of dinner overview. How the UK Biobank which is the biggest biobank and prospective. Successful so we'll see.

07:30

Good morning, everyone.

07:33

Today, we're back. So earlier in the week, myself, I've already talked a little bit more about the general context and rationale. Why don't you just at all. Today we're going to do more of a deeper dive into what data is why we have that data, how it looks in the morning and how you can access it yourselves. And then in the afternoon, late morning affinity symbol, kind of demonstrations of how you can use the data different ways you can interact with the data. Some examples of code and we have some training resources which you can share as well. So for anyone who has access to the data or wants to try them on their own resources, you'll be able to get access to those and make your way through them in your own time as well. So a start of an overview of UK Biobank just catching everyone back up so.

08:58

It's okay. So you get by biobank recruited 502,000 people aged 40 to 69 in 2006 2010. About 90% of these came from England 7% from Scotland and Wales. And these were recruited at 22 assessment centres which were distributed around the country. So from north up in Edinburgh, in Scotland down to Croydon, which is a suburb of South London, it's not very nice. So invitations are sent to those meeting a criteria who lives within a short distance assessment centers. So in general, it was people who live about 25 miles 40 kilometers from an assessment center or backwards. But from the NHS, if not public health system through general practitioners so we can get the data for those of people who lived in the Barrett Jackson's in those areas that the age criteria and then received an invitation by post. Now these locations were chosen for access, the diversity of population, working population density. across the UK, you have different ethnic breakdowns, different age breakdowns, and also these very fairly rural areas. So Wrexham in north Wales is very small, down to three which cover central London which at the time had about five or 6 million people. And during recruitment about five and a half percent of people who received an invitation in the post to join. So it's important to point out that this centers on balance in terms of size, so the smallest center is maximum, and that only recruited 650 participants, but they're important participants to get the kind of productive diversity, whereas Lee Leeds is the largest and that had 44,000 participants, so almost 10% of the cohort, so a lot of the ones in large cities. So Leeds Birmingham had sort of 30 to 40,000 participants in the single center and then 20,000 participants. Manchester centre will also ask to come back for repeated baseline assessment in 2013. So what did they actually do when they turned up so baseline assessment it was also an in person visit to us some top participants were asked to run it had details of their family doctor. The medications they took, undertaken the past operations as hard and approximate date. So people number one, any medical history they knew about their family. So diseases that parents or siblings. I don't think they knew about their own birth as well. So we report things like participant birth weight a lot of people don't know. But for those who didn't know, it was useful to have during the initial stages in the center as well. They went through a consent to their cards, insert a CD with that invitation I have detailed information about what they would be consenting to. That was then talk through them through with them by wanting to start in the center and see what happens. You can see they consented to share all their medical records and any other health related records. And they consented that they could be contacted for further assessment in addition to blood samples and so then the first part was and this is where an awful lot of the tabular data which UK Biobank has comes from. So this took participants about 15 minutes the filler and it went through lifestyle questions. So you can see there's an example on the picture on the board. So that's what kind of accommodation do you learn to do live in a house to live in? A Tahoe? Do you live in an apartment? They have socio demographics. Early life history probably grew up psychosocial factors, medical conditions and medications and they remembered and their family medical history then translate and put that into the customer. And then there was some set specific questions and female participants are asked about their previous pregnancies about if they took contraception male participants were asked about some other related conditions. The age they've gone through puberty, things like that. While I was still using the touchscreen device, participants also went through a pairing test. And so this is speech and noise paths developed by a university in the UK for UK Biobank. And this is where the participant identifies three numbers which are spoken quite quickly together with a lot of background noise at the same time to identify participants can get meaningful information from a loud or busy environment. They did cognitive function tasks so if you've ever done a selective entrance exam for or aptitude testing probably done something quite similar. So this is verbal and numeric reasoning, but that also their memory and their reaction time. And then from 2009, we included 24 hour diet to recall this is Oxford which is used by a lot of our studies, one out of the University of Oxford human study, which covers about 200 foods and drinks. So how many portions Did you eat in the previous day of bread? Or pasta or apples? Any supplements and dietary choices to do they follow a special scripted diet and one of the major choices you have I made was to do all of this through touchscreen interface. So at the time, one of the most common ways to do these kinds of records was through paper questionnaires, or only sort of verbal and today that wasn't possible to do it all with verbal interview with the number of participants we're seeing in a day to ensure we can get people from into centers swiftly and efficiently. Recording on paper is messy. There's records it's hard to understand writing people still have to do that. And so this is kind of a choice which was piloted to see how well participants would react to using a touchscreen interface and was well received and so he's throughout the entire survey. So after the touchscreen questionnaire, the next step was a verbal interview. So there was specially trained assessors who took some of the information the participants and put on the questionnaire and ask them about their medical history. So if they said on the questionnaire that they previously had a medical condition, asking for more details so they could get up a specific coding to understand what that condition was sort of participants as I have a heart condition, you know, do they know more about what that is? So you don't need to have an awful lot of drop downs within the same scenario which might make it complicated for participants to use, but someone who is familiar with medical terms, right. And they also asked about the cancer and the operation. And they typically about customers numbers from 2000 to nine division acuity auto refraction and intraocular pressure, along with an optical coherence tomography scan. So these are similar to the kinds of tasks you might get autism monitors to look at the heart, eyes, participant reason and the physical measures. So this example in the picture is someone using a dynamometer. So that measures how strongly someone can correct with our hands, which you did progress, the left and the right times. There's standing and sitting heights or sitting down on a chair. I'm taught my work from the waist up but also how it all went standard weight and bioimpedance are typically symptoms that are presented wise after the pilot that was introduced, which is a way of art, bone health, and also spirometry. 78,000 participants got the cardio measures which included a 12 lead ECG both at rest and during exercise everyone who was healthy enough to do it.

18:30

And then the one minute ECG article covers the exercise challenge on our fitness side and see how well that performs. And then the last part of the assessment center was to take a lot of blood samples. So 85,000 participants we took a saliva sample. From all participants we took urine and blood, which was split into various different tubes and then within the TTI that was also split from Buffy extracted DNA samples as well. So these were taken every day they were they were extracted into the remote centers, and then sent overnight to our main arteries which the fire going to be answer them. So how does this recruited population look like compared to the UK as a whole? So we've created 54% female participants, and that was slightly higher recruitment and old age group. So we've had described in the agency joining over the top and that was higher tape. So the median age of recruitment 94.6% of the cohort to find themselves ethnically as so this is fairly similar to the UK census in 2001, which was just reported pregnant. It's a bit less than the UK census in 2011, which is just after recruitment for us. So it's slightly higher but not drastically higher than the UK as a whole. As it sits now. It's quite a population. So there's been a lot of growth and ethnic diversity over the last 20 years. So our population as it sits now is a lot lower percentage wise than it was at the time we recruited. But this was a reasonably accurate snapshot of the population at the time, particularly in that age group. One thing which surprised us was quite the number of relatives recruited in the cohort. So you can see from the graph you have some participants, quite a large number of participants who have over 21 relatives who are recruited along with them. Even today, we encourage people when they come to clinic visits to come with any other family members who are able to come because people quite like that. It was a day out. They go together, you know they travel together. They have to have an overnight stay that's done together. And yeah, so we do have an awful lot of participants who have relationships and from that we also have about 1000 trios, so far father and child repaints. In the cohort, the health of the population has a similar rate of cancer diagnosis to the UK population as a whole however, if you break this down into which cancer diagnosis they have those which are affected lifestyle. So for instance, lung cancer can be affected by smoking liver cancer by alcohol consumption. These tend to be quite a bit lower than you can population averages. And those weights are picked up some routine screening measures. So breast cancer and cervical cancer, these tend to be slightly higher and identified slightly. So they make up a group of course, and all cause mortality rates in the cohort are slightly lower than the effect as a whole.

22:23

So since 2014 participants have also been invited to come up for a comprehensive imaging assessment, which looks at the MRI scan using three tests and brain, abdomen, liver, kidney and pancreas in detail. They also have a DEXA scan so that looks like bone density decomposition. And then we have about six and a half 1000 participants who come back for pizza. So this is done about four years after their first scan to the development of any conditions they have to increase the number of deaths as they look towards aging, particularly in the brain areas. I mean, oh CT scan was also performed during the second half of baseline. So for some participants will have three or four images from the imaging. So these are given as more editing files and those are made available to all researchers, but there's also a large number of groups who create equal imaging derived phenotypes. So they use the data pipelines to break down those imaging types to bring out specific pieces of tabular data, which makes those images more accessible to the researcher population. as a whole. So for instance, the cardiac is looked up by Imperial College London and Queen Mary University of London. And they work out the dimensionality of all the areas of the park and the images return that to us. The brain by Pembroke and access and this is one they've got the volume of white matter. And then hundreds of discrete pieces of information in those parameters. As a group and Southampton can do similar for the DEXA data and then perspective and Uppsala, consider the dominant organs and have a look at the body composition information to us. And then the number of visits per participant, if you consider as well to see how much repeat data there is. So about 80% of the cohort at the moment has only attended one visit is nearly 100,000 to two. So that's either baseline baseline or baseline and why we're doing that through three parties. Three visits are by about 20% of 20,000 participants and then there's a very small number at the moment. You have all four visits. And the interval between visits can be a small number of years, which is particularly common to those who have three or four minutes whatever time you have every few years, up to 6050 years. Those who attended may be baseline early on and have now recently come into the imaging study. So one of our monthly conservative participants for was online. So for us to be able to make contact with them. And the main reason we do this is for questionnaires. So we send out one to two questionnaires. per year by email to all eligible participants. Participants can then choose to respond to that either as a web questionnaire or to phone us up and go through the questionnaire details with members of our participant call center. So that each cross dinner is covering a different discrete area and is developed by a panel of apps. And these kinds of be a combination of validated clinical or research questionnaires and custom content. So for instance, the mental wellbeing questionnaire in 2022 included some common diagnostic questionnaires is the diagnosis of depression or anxiety as part of the class students Dawson general questions about well being that were considered by our experts to be a really important addition to the resource. So there's cover a really wide range of pieces of information. diet has been studied quite extensively, looking at sort of diets and also food preferences. And for a number of participants as well. We have they just have seasonal effects of their diets and just enough so as to go over it multiple times. And then there's more upcoming questionnaires as well. Looking at social interaction and focus and then across the questionnaires. The 24 hour diets questionnaire is by far the best responded to it because it was included as our baseline assessment and the last sort of 20% of participants but just over 10% of the cohort have responded to every customer. So the thoroughness of that data is quite rich. There's a lot of participants for which you can get every single piece of information across multiple years. And then there's also some additional follow up which has been done either as a add on to the visits participants have in clinic or by post. So accelerometry was carried out in 2013 to 2015 and participants received these values contain certain risks one accelerometers, so that then looked at that movement for seven days. It was a blinded device, so there was no screen telling them the number of steps they've done or that heart rate. And this recorded information as a sort of discrete tri axial movement, which can be translated into describing how someone likely how they were active if they were asleep. You can see this person likely went for a jog. They had some time sedentary during the day and time to sleep. For about 25,000 of those participants received the accelerometer up to four more times through the year. To see if people's movement varied between summer and winter, how active they were through the year. And then participants who have attended the emogene clinic also are over 65 years old, are asked to wear a cardiac monitor for the two weeks after the eighth. So this is a device which is kind of tasked with medical type which records the heart activity and it also includes an accelerometer, so the activity of the heart can be compared to the amount of exercise the amount of movement like during the time. So that's an ongoing study. Data collection is expected to wait about another 12 months so I expect the data will be available towards the end of 2020 sets the right 2020 sets eventually be available to all resources. And then these biological samples which we gathered also been translated into a large number of different types of data. So starting with biochemistry and infectious disease, and genotyping. Unlike proteomics, there's more studies upcoming so the proteomics data was a targeted funnel with 60,000 participants, or 3000 proteins concern this is a explorer originally about 1500 protein panel and then expanded the phrase housing person and this included samples from the COVID energy. So during covertly prioritized visits for participants who have been identified as having public or private and not controls to those participants in my raffle depended on PacSun to ensure that we could provide some replacements a little bit of physical changes biological sample changes, changes in mood and lifestyle between us. So this data that are causing inflammation, urology and oncology offers over diverse individuals and it includes a mix of randomly chosen participants. These COVID-19 imaging participants and also a number who was selected by the Consortium for having specific diseases of interest about research.

32:18

And then the maximum load assay carried out with mitochondrial health who are analyzing 239 biomarkers because in samples, there's probably 290,000 samples of data available, and that will include a full cohort at the end of this year and in our study, they're looking at both those baseline samples on the first date to changes over time for both participants. 15,000 samples from us identified 800 new life and protein medical disorder diagnoses that got 700 hypotension diagnoses in that interval between the first and the second sample and these participants include over 600 ICD 10 codes and the conditions okay, so that's an introduction to the data path, and now a little bit about why that is and how to get it. So the best way to browse the data is on the website we call showcase UK Biobank data which is available online and the information and slides is public or receptors. You don't need to be signed up to an account or member in any way to access this information. It's freely available about the data fields. So this page includes various kinds of top level categories. So racing and grouse concerts and looking at the catalog information there's some downloads of supporting utilities. Essential Information is well, I would strongly encourage you to go first if you're interested in looking at this data. So this includes really detailed information about the surplus resource but also about the UK Biobank study about how all the different visits were created, scientific and statistical information about the cohort and how the cohort was selected, how each element of the study was designed. And there's also information about how you access to data and returning results.

34:57

So if you want to look at the data fields which are available, the easiest way is to browse if you're just curious as to what that is. So this allows you to see the data was nested categories and subcategories, generally organized by data source. So if something is from an online follow up, so that's a questionnaire and something was gathered in the assessment center, or from a biological sample. The items that that indicates the number of disparate data fields within each category. So you can see that there's various imaging we have nearly 3000 different fields different types and pieces of data gathered from that. While some of these like biologically sound that will be mostly procedural metrics and other samples. So if you've chosen the data you're interested in and you want to have a look at it. You can open this up from that nest. There's nested categories. You get this kind of summary information about the field. So you'll have the field ID and name so this is albumin. It's from the nightingale metabolomic study, and then tells us how many participants are publicly available. So this was 275,000 participants and then how many pieces of data so because each participant can have more than one answer for each measure, you can say that as about 15,000 to half a second as not integrated by the other number. There's always a summary of the data range provided so that fits the minimum and maximum the breakdown within games and media. And that's graph as well. It provides information about the units so this is continuous data, grams per liter and when that data was initially made, made, available, which is February 2021, and when it's been last updated. So we've had about 100,000 participants other diverse data, and to say that, but then there's Costa which I think days is going to smoke, our next presentation. It's really important to know about it, not even trying to guess this looks laid out. And then the other thing tells you is how many instances as pass. So data is split between instances which are essentially time points. So now you can see as to so incident zero is the initial assessment, that's the recruitment process. And then there's incident one visit here, which is the first assessment for those participants who came up with baseline. So the instancing is typically the occasions where it might not be for instance, with dietary it was all performed online and they didn't have to try to visit the participant and then that's the time to personnels now is a schema which provides meaning instances for the field. So they are provided in the downloads page on shared place, which explain how the data fits together. So if you've done this, it was that, you know, that field was great. They recorded as the field ID, das, the instance. One for instance, for the person to keep assessment dot zero, which is the right and then on rap. Importantly it's shown slightly differently. So uses P incontinent field and then i in front of an instance and they arise left off if the data isn't the right providers if there's multiple values measured at the same time. So if I take blood pressure three times in a row during baseline hours of one instance, it's all done on the same day, but like the numbering all starts at zero because I chop off tech works in C++. And the route moves off an array if it's not used for the field. And then there's a difference for data fields which are categorical data. So now instead of seeing a histogram, you will see the data split by each discrete item within that and it also links to a specific date to coding, which tells you what the data means. So when you download it instead of January, it's important in the data, and you need to use that in combination with a data codec to understand how you translate that to the downgrade and then the day to summarize that is a bar chart as well within each field as a Resources tab, and that provides documentation sometimes code how incessantly file and that helps you understand the data that it's dumped in many fields and categories and it can be downloaded or accessed by anyone without any registration. It often provides context to how data was collected and quality control. I think if the fields you're looking at are key parts of your research question, and you're considering using the data and physics specific fields which are available, it's really essential reading to have a look at how that data was gathered. So for instance, this is on the blood biochemistry we have two resources now which tell you about the quality procedures. see information about the biomarker data. That then would provide you information about how the assays have selected the methods used instruments daily and monthly credit quality control and verification, the quality assurance program so if you're looking down the line to publish your Miss data, it's really important that you understand all these elements about how it was downloaded about potential strengths and weaknesses of the data. So you can ensure that it's built into your research process. And then the other thing which is useful within the cybercafe and that has information on related field, which can provide important contacts to the data field which is provided. So here this is the abdomen field, and it has a related field, the TCP flag format, you can see and that would then give you information as to whether there's any known issues which might make the measurement less reliable. So if it was below the limit of quantification, if there was a flag that there was potential contamination within that measure, and you would need to consider as part of your research when you match to these fields, if you want to split this data or use different normalization. So I'm just going to like quickly talk about a couple of things that you might want to consider when you're using this vector. So we have an incredibly large data set. And that means that lots of things have been measured slightly differently, or many times over the years. So for instance, if we've taken a participant height, floor times might measure slightly differently, partly due to instrumentation or letters or changing heights. That sometimes I'm like hypertension, so high blood pressure, where multiple data sources can lead to either inconsistencies or different ways of interpreting the data. And you really need to consider and clearly define the private area in your own research. So hey, we can say for hypertension, we could do that as participants who've told us when they come to the clinic, and we asked what health conditions do you have they tell the person on stage hypertension, that's an easy way to to define that. Then we have a blood pressure measure we take. So then there's some of these participants who they have been recorded as high blood pressure, but they haven't told us that. So did they have white coat syndrome? That stress in the clinic and our blood pressure's high? Do they not know they have hypertension? So they haven't told us but we can see that they could be diagnosed. And then you can look as well at medical records and things coming from a GP and you can see that there's about 40,000 participants who have been diagnosed with hypertension, but they haven't told us they've come to the clinic. So maybe they've forgotten. They didn't write down all diagnoses or they don't really consider it a problem, or they didn't want to say when you can look across the different ways and you can see, there's different ways to determine what you consider your criteria for your research. And it's really important to think about how you would do that and make sure that you're consistent within your own analysis.

44:44

And then another example would be insulin dependent diabetes. So if you look at participants who have a diagnosis of insulin dependent diabetes mellitus, that's an ICD 10 code and at the time they attended baseline, who are taking insulin products, there's 793 participants for whom we can see that there have been prescribed insulin and they've told us that they have diabetes, because you want this 41 participants who have a prescription for insulin, we can say that they've got a prescription for insulin, but they haven't told us that. So they haven't told us either that they take it or that they have a diagnosis of diabetes. Now with some medical conditions, that's about harder, there's lots of reasons you might take the same medication. And so then there's not so many reasons to perceive it. And then you also have participants who are self reported that they have diabetes but we don't have a record that I've been misled into. And then if you look at the total number of participants who have any time diagnosis, that will be about 5000. So there's about another two and a half 1000 Who we have no record of the incident. They didn't tell us there's no prescription data. But we can see from GP records from often our hospital records, often their death reports that they still have this diagnosis. So often when you're using these kinds of complex datasets, there isn't an easy way to dive in to determine the criteria for the health conditions which are interested in. I'm really considering all the sources of the information, how that might overlap, where you might get differences between them and why making sure that as a researcher don't make a choice as to how you define those conditions is a really important first step in using that.

46:58

So we'll leave that because that is an awful lot of information.

47:04

And I think there is a I'm happy to take questions now. Or we can do questions at the end, whichever suits any questions along the way. Oh, yes. I don't know if there's any.

47:26

Thank you so much for the beautiful presentation. Gentlemen, Rusty. I'm following up. Japanese 3000 cohort of pediatric intractable disease. With blood loss it is as poor as more us than I see you are Hezekiah hot. I think it is not likely to get that increase like a response, increase response like your cohort. We are suffering from embracing the less bonds lead from the parents, especially if the students are dead. So tips for increase the response rate on the subject disease covered in pediatric patients,

48:13

I think, yeah, I think what you're doing is really hard I guess because it's as well, you have one of the advantages of our data sets is it's adults who are consenting for themselves. So the person who is in the data set is actively keen to be part of that. And they're doing it without any kind of health condition, specifically as the process. So there's no kind of period of interest which might drop off if that condition dissolves, or they feel that they no longer want to be involved in that research. I would say response rates are hard. We got really good recruitment. We still get very good uptake towards imaging doesn't suppress the last, but it does do a heartbeat participant if people become less able to take care of themselves to travel on the road. I think it's always going to be a really big challenge that these kind of long term studies I think will be even more so of your recruiter recruiting parents and children. Okay, thank you so,

49:26

I have a question about omics. There's too many more from tuba pharmaceutical. And so as a pharmaceutical researcher, so we have recently started using recently published data set and I feel it's quite magnificent. For finding new biomarkers. And if I remember correctly in the presentation of data, so someone introduced in future that you get around to the extent that block omics datasets and so it is explained that so, so, you be a collect blood or coding data set from 100. So

50:14

we are in the process of doing 50,000 participants. Proteomics, so there'll be about 10% overlap because I like to increase the number of data to around 100,000 And that project may increase past 32. And there is a lot of interest in increasing target targeted proteomics. So for instance, the opening of potentially a different one to four cohort, that's not yet being informed as to the timescale of it. But I think you're quite confident that in the next key is not present.

51:05

Your presentation to us to do something about constricting and those sorts of things. So, but you know

51:18

tonight's presentation after the break, so yeah, we have

51:24

Yeah, I'm looking forward to

51:32

Minnesota, we don't have enough room to stand up. So, brings you to wonder will expand to them, we only have a mass spectrometry proteomics data, you can specify specific part of his films or were targeted in some way based off of their characteristics because it would be really interesting if you were to consider ICD 10 codes and the proteomic profiles of controls for certain diseases.

51:59

Yeah, so if the original 60,000 participants for 10% of those were specifically selected by the consortium members who were funding the study, to the diseases which are of interest, cardio, embolic,

52:15

or flight, I forget what the four categories are, but those those

52:21

are the categories all three sorry, I didn't explain that one. Well, those are the categories of proteins. We try to target. The selection was dependent between each consulting manifesto it was agreed across, there were 12 Consulting members who chose sort of destinies ventures and worked together to determine so they vary from you know, having people have specific ages when they attended for from Congo attended before after certain diagnoses are part of the data which has already been gathered so maybe they already have everything data set. So the reasons those participants were chosen in various ways, and so team members research interests with the mass spectrometer for whatever x you are doing random, completely random 50,000 There'll be 10% overlap with OLED but that's just to kind of enable normalization between the two requires. But there's no specific selection criteria on those participants. I think the targeted proteomics happened

53:35

before

53:38

so we went through all selection. Eventually, generally trying to if you will have people who are interested in particular, especially fairly brand new persons, participants are all within a single batch quality control within that batch. So how that might affect normalization. You heard it works in the data, as opposed to just having to sit tight and up and down until all your participants.

54:17

presentation. Thank you. My question is about the severity visit. So you mentioned that the repeat assessments were done in Manchester. Yes. So does that mean that only the participants just for the first time could come back? Or simply

54:36

have one clinic for that which was in our headquarters in Manchester? That stopped because we have the funding for the imaging study, and that became our pilot activity center instead. It was based on the drastic participants so again, anyone within 30 to 40 kilometer radius of the Manchester a lot of people won't have moved, so they will probably have gone too far their initial recruitment to Manchester or perhaps one of the nearby cities like Leeds that they didn't have to look at the answers.

55:20

You should also have like, at least worked with the participants as a whole. They were quite representing their population but like for the population, is it mostly white groups?

55:33

I'm not sure but I know there was a specific in that sentiment Mason category. I published guys a really detailed about the 15 page document which was done by our epidemiology. Team at the time, which looked at real state of Montana, about what the baseline study how that compares to the cohorts as a whole and how that compared to the population as a whole. So that's definitely now I know, Manchester is one of the more diverse areas. So I would expect that there isn't a huge change and ethnicity. I know for instance, particularly with our baseline, some of our peak minutes, it's harder to get no other participants who tend to still be working overtime to take time out of work. souring responsibilities, younger children, as far as a finance director compared to people who are retired or working part time.

56:39

Maybe I can squeeze the last question into the next section that was curious about the yearly questionnaire. Assuming you're getting a lot of demands, and the wiki even though we also don't want to make this too too long, so yeah, what's the process that we went through to select questions?

56:59

There is a whole team who are about six people whose job is trying to make everyone happy when it comes to questionnaires. It's not an easy one. So part of this comes through requests often people are messaging across to contact participants on an application. And that might be converted into a counseling question, as opposed to explicitly contacted participants. So we get interested groups and then to ensure that you have a full like customer say someone is interested in doing a specific metric for depression. You might read that into a mental health questionnaire or generally by the other leaders in those fields and suggest questions for that. And then there's quite a lot of trial and error. A lot of childhood stuff of just as the questionnaire legible. We've had one on facial recognition, which is all in so it was could you recognize celebrities with just that face, and that's hard. And then they're tested on a small panel. So randomly chosen usually around 500 to 1000. Just to see response rate and how participants find that goes well. So are you