CENTRE FOR LONGITUDINAL STUDIES

COVID-19 surveys in five national longitudinal studies, 17 June 2021 transcript

Richard Silverwood 02:33
Great, thank you very much. So hello, everyone, welcome to this workshop on the COVID-19 survey data that we have across five national longitudinal studies. So we're really excited about having this COVID-19 survey data. So it's great that we're able to share our enthusiasm and hopefully some information about it with you today. First of all, a few standard housekeeping notices. Please do keep your cameras off and your mics muted at all times. I think you might be able to see that this has been recorded and it will be made available subsequently on our website. So just so that you bear that in mind, if you have a question, please use the chat function. And do note that your question will be visible to all attendees. If you've got any technical issues, then if you're still able to access the chat, then you can use that, otherwise, do email us at our CLS events address. And towards the end of this session, we'll post a link in the chat to a short survey that we have. And it's really useful when we're running these sessions if we can get your feedback on what's worked well, what hasn't, and the sort of thing you'd like to see us running in the future. And as I say, thank you for joining us today. So what are we going to be looking at over the next hour and three quarters or so. So first of all, I'm going to give a very brief introduction to the cohorts and to the COVID-19 surveys, then Kate's going to go into some detail about the content that we have in the surveys across the three waves. I'll be back again then talking about some approaches to handling missing data within the survey data. Yes, within the COVID-19 survey data. And then following that, we're going to have two slots, one from Vanessa, one from Bozena, and with some research examples using the COVID-19 survey data. And these both use data up to and including wave three, these data have only become available recently. So it's really exciting that we can have some very fresh examples of research that's been undertaken using these data. Then finally, we'll have a Q&A session. So I'd invite you to ask any questions that are relevant to specific sessions on the way through, we'll try and address those at the end of each session. But if there's any more general questions or anything left over anything we've not had time for and then we'll cover that in the Q&A session at the end. Okay, so as I said, I'm going to give a brief introduction I should introduce myself. I'm Richard Silverwood. I'm an Associate Professor of Statistics and also the Chief Statistician at the Centre for Longitudinal Studies at UCL. So, the COVID-19 surveys that we've conducted are across five national longitudinal studies. And this is just showing those and the ages or sorry, the time ranges that they cover here. So the first, the National Study of Health and Development (NSHD) is

actually run by our colleagues at the MRC Unit for Lifelong Health and Ageing at UCL. The remaining four cohorts are the National Child Development Study (NCDS), the 1970 British Cohort study (BCS70), Next Steps and the Millennium Cohort Study are run by ourselves within the Centre for Longitudinal Studies at UCL. So occasionally there is a bit of differences in terms of, for example, data access, as we'll see shortly between the CLS cohorts before at the bottom, and NSHD. So one bullet point each on each of these cohort studies, which is clearly nowhere near enough to be able to describe each of them. But just as a very, very brief introduction, I'm assuming by registering on this that many of you will be very familiar with these cohort studies anyway. So all five of them are ongoing longitudinal cohort studies with a multidisciplinary content spanning areas such as physical and educational development, economic circumstances, employment, family life, health and wellbeing, health behaviours, social participation, and constituents. NSHD comprises babies born in Great Britain in one week in 1946, with an initial sample size of over 5000. The NCDS is babies born in Great Britain in one week in 1958, with an initial 17,000 or so that were later augmented with some immigrants. The BCS70 is babies born in the UK in one week in 1970, again, with an initial sample size of just over 17,000. Although Northern Ireland members are not followed up after birth. More recent studies have slightly more complex survey design, each being more a cluster design. So in Next Steps, rather than being a birth cohort, this is people living in England born in 1989, or 1990. But the study only began in 2004 when they were in year nine at school. So as I say, it's a cluster design that was sampled at that age, age 13 to 14, with nearly 16,000, in the initial sample size, with the latest booster sample of ethnic minorities. And the MCS, as the name suggests, children born in the UK around the Millennium 2000 to 2002, with a cluster design and an initial sample size of almost 19,000, with some later additions. So I think this is a nice plot of the study timelines, showing how the cohort members sort of track over age and over time. So for example, the purple arrow shows that members of NSHD were born here in 1946, at the bottom, progressed over time, so that by the time we get to 2020, they're in their 70s. And the markers show the points in time, which we make observations from the cohort members. So we can see that they're slightly more frequent in childhood, but still at regular intervals through adulthood. Similar sort of pattern across, excuse me, across each of the other cohort studies, with the exception of Next Steps, which is the same we didn't sample until age 13/14. So there's no observations prior to that observed every year after that for a period and then more recently. So we've got these five, long running very well followed, very well characterised UK longitudinal studies. 2020, the pandemic hits, and it gives us a, you know, an opportunity having these well established longitudinal studies to try and do some important work around COVID-19. To meet these ends, we introduced a set of three COVID-19 surveys, so three waves of COVID-19 surveys. The overall aim here was to understand the economic, social and health impacts of COVID-19 or COVID-19 crisis more generally, and the extent to which this is widening or narrowing inequalities, and the lifelong factors which shape vulnerability and resilience to its effects. So because we've got these data going back many years in some cases, decades then we're in a good position to be able to address some of these questions. So, in 2020, when the pandemic hit, these are the ages of the cohort members in each of the cohorts. So we can see that they vary between 19 in MCS and 74 in NSHD. So there's a good amount of variability there over several decades of age, which allows us to use our COVID-19 survey data to think quite a lot about cohort and age differences in a fairly systematic way. So when exactly did we conduct these surveys? The first wave was conducted in May 2020 and that was really at the height of the first lockdown. The second wave we issued in September and October of 2020. So at this point, and there were the you know, eased restrictions, but because of the nature of the survey, we're looking back in

time to a large extent and thinking about changes since that first wave. So it's, it's really covering all of that period in between waves one and two. And then the third wave, which has been recently completed in February and March of this year, that was during the third lockdown. The mode of the data collection was web only in waves one and two. But in wave three, there was also a subsequent telephone aspect to it whereby non respondents to the initial request for response by web were followed up by telephone. So a lot to digest here, but just a kind of an overview of the response set out by cohort, and separately by each wave. So we can see that within each wave, the extent of response, so this is response within the issued sample is quite well patterned by cohort and these. So with the exception of MCs parents are a bit anomalous, the cohorts themselves, the cohort members are ordered by age. So we can see that in our older cohorts, the response has been much higher than in the younger cohorts, we see that that same pattern is the case across all three waves. We can also see that our response has generally been increasing within a given cohort across these waves. So we've been improving our response rate. And you'll notice that the issued sample itself jumps quite a lot from the first wave to waves two and three. And that's because in the first wave, we were issuing this obviously, quite short notice, because of the pandemic, we didn't have the capacity to issue to sort of mass mail out. So we were restricted to issuing our survey to cohort members for whom we had an email address. And subsequently in waves two and three, we were able to do so by post so the issued sample increased. So a little note on accessing the data. So, data from these longitudinal studies are generally available via the UK Data Service. And this is the case for the COVID-19 survey data. So the CLS cohorts - so that's the cohorts excluding NSHD, are all available via a Standard End User License here and under this study number, and downloading this will give you data from all three waves of the COVID-19 survey. NSHD data are available also from the UK data service, but separately under a slightly different licence agreement. If you're used to downloading data from the UK Data Service, then this is the standard documentation list that comes with the data that are available for download. And this is the grab for the COVID for the CLS COVID-19 surveys, and it's got the standard things that you'll be probably used to seeing. So variable lookups, and data dictionaries, copies of all the questionnaire so that you can refer back to exactly what was asked, and technical report. And I think probably the most useful for most purposes will be the User Guide that we've put together, which I'll show you in a second just briefly. And these resources are all available from the UKDS website, but also from the CLS website where we have a dedicated COVID-19 survey page here. As I say, the User Guide's probably the most kind of useful or understandable of those in most settings. And this is it here, and indeed most of the information that we're giving this afternoon, certainly across the first couple of sessions can be found in much more detail in the User Guide. And I'm sure at several points will be saying, you know, for more details refer to the User Guide. All I plan to say about any questions just on the introduction before we move on to more kind of substantial content? No one's given us any questions yet. So great. I shall take that as a no, thank you. Great, hopefully, I stopped sharing that and I will pass on to Kate. Thank you.

Kate Smith 15:56
Okay, thanks very much, Richard. Just waiting for my presentation to load up. Okay, so, my name is Kate Smith, and I'm a Survey Manager at CLS. And I currently work on the Millennium Cohort Study. But I have worked on all the projects we look after except sat Next Steps. And I was also part of the survey management team, which put together their questionnaires for the COVID surveys. So really, I've got some some things which summarise what Richard is just said which we did three online

surveys with all five birth cohorts plus the parents of MCS during the COVID-19 outbreaks and wave one of the fields work was carried out internally by CLS given the very tight timetables and the sort of consequences of lockdown, but waves two and three were carried out by Kantar Public. And as Richard said, the aim of the waves was to collect insights into the lives of all of the five cohort studies participants at the height of the lockdown of wave one, sorry, was to look at the effects of the lockdown restrictions in place in May 2020. And wave two, we aim to capture how those lives have changed since that lockdown and through the summer, last year when there was some easing. And also, most importantly, there was the return to schools and education in the late summer and early autumn of 2020. And for the third wave, which we just wrapped up a couple of months ago, was done at a time after the new renewed restrictions were imposed post Christmas 2020. And as Richard said, non-responders to the wave three online survey were followed up by telephone. So we aimed for the most part to harmonise the questions across the five cohorts, although there were some cohort specific questions, areas and also we wanted to maximise the use of longitudinal measures. So all of that rich data which we've already collected, individually about the five cohorts. So for the first worth, first wave of the data collection, which was really the marker, and set the sort of picture of things as how they were last May, one of the most important areas was obviously to look at physical health and to collect data from questions on whether participants have been suffering from COVID-19 symptoms, whether they'd had any testing, but also to contextualise that, alongside their general health, and also to collect any information on any health conditions that they were suffering. We also really wanted to look at mental health situation and specifically loneliness, I think we were already aware in May and during the early months, when we were designing these questionnaires that we were really in a in a very unusual situation, which none of us in living memory can remember living under. And so we really wanted to look at the mental health status of the participants and also elements of loneliness, and whether they had access to social support, but also get some measures of whether there was conflicts being caused by the situation that people found themselves in. And also within the context of health we also looked at healthcare and receipt of care. So whether people were giving or receiving any healthcare of any care, not just health care, but any care during this period. We then went on to collect some demographics about people, their housing circumstances. A very, very short household grid, nothing like the normal things that we would do in one of our major surveys but a real just a marker of who, who people were living with. But we also want find out whether there have been any changes to the living arrangements caused by the pandemic either people coming, returning home who perhaps wouldn't have done so otherwise, and also partnership status. We're also obviously very interested in the financial impact of the pandemic at that wave one and labour market outcomes. So whether people were managing to maintain their work situation or whether they had experienced furlough, or whether they they weren't being able to work at all. And obviously, very importantly, particularly for the MCS, MCS families. So there's not so much the MCS cohort members who were obviously only 19. But the MCS parents, so if they had younger siblings, but also the MCS, actually, in terms of their university, if they were had embarked on university careers, their disruption to any education and for parents, their experience of homeschooling and how that was panning out in that crazy chaotic time. For MCS, we did ask about their experience of university education during the pandemic. So whether they were being able to continue lessons, and lectures online, and also how they were feeling about their experience. So if if any of them when their first or second year, this certainly was not the university experience they had signed up for or were expecting. In wave one, we asked some time use questions as well, because there was interest in how people were managing to fill their time or the sort of activities they were

undertaking during this period. And then we also asked some questions on attitude. So how trust in government was panning out whether they had support from their local community, and also markers of risk and patience. And then for everybody, we asked whether they would give their permission to link to the Zoe COVID app, the King's College COVID symptom tracker app, and whether they would be willing to link to that so that the Zoe app could track their symptoms going forward. And the opt out rates for that for wave one, we're about 27 to 39%, depending on the cohort. So wave two, when we went back in September, October, I mean, to go back to the cohort members, we mainly repeated the questions that we'd asked from wave one. So all of the questions about their experience of whether they'd had COVID symptoms, whether they'd been diagnosed their health markers, their mental health, all of those we repeated, but particularly what we did for wave two was we wanted to focus on parents and children's education, we developed a child loop. So for anybody who had a child aged five, to 18, in the household, we asked them about what had happened to their child's schooling during the summer term of 2020. So these included whether there had been physical attendance during that period, all of that period, or some of that period, or not at all, how home learning had gone, whether they were able to be taught via online learning or not, because we're very aware that different schools were able to do different things, their online lesson provision, also where the parents had had to step up and step in and become teachers. And that was also obviously both mothers and fathers and also the availability of learning resources. So the availability of separate PCs, or tablets, or some form of internet resource, whether the kids had some way quiet or dedicated, that they could learn at. And also sort of questions along that. We also asked the parents their own assessment of how the child's academic progress had been affected, and also their mental health. So prior to the COVID pandemic, but also currently as well. We improved some questions on access to health care for that period and we also asked new questions about life events. So we asked sort of life events in key areas and the year before the lockdown and then since lockdown. And wave three, the most current wave of the data collection, obviously, again, we mainly repeated questions from wave one and wave two, so the content was largely fixed. However, there were some significant new areas for wave three, most importantly, I think the vaccination programme which began to be rolled out, sort Christmas time/January. And so we've asked we asked all the cohort members whether they had had a vaccine at that point, and if not, whether they intended to have a vaccine when offered. We also introduce some questions about their experience of Long COVID. So for those cohort members who had experienced felt that they experienced COVID or in fact not, not as well of suspected COVID we asked them how to arrange of questions, which tried to tap into whether they were experiencing Long COVID. And we also introduce some questions about whether their compliance with social distancing guidelines, and also whether they had downloaded the NHS track and trace app as well. In addition, on the financial side, we asked, put in additional questions on pay and household income, and also the amounts of financial help that they might have given or received. I should repeat that for both wave two and wave three, we repeated the questions, asking them to if they had not been able, at a previous wave to link, give their permission to link to the Zoe tracker app, then we asked them again, whether they would do that. So I having given you an overall I mean, I do hope that you can see this screen. So this gives you a much more detailed account. And I'm not intending in any way to go through every section. But within the physical health, it gives you an idea of the all of the questioning and which waves these questions appeared at, so you can see that the for instance, the Long COVID symptoms are only asked at wave three, the extent of implies for social distinct only asked at waves. Well, we actually asked some questions at wave one, we didn't ask them in wave two. But we asked extra questions at wave three, and the NHS track and trace

app again, only wave three. And we asked a pertinent question probably, how well the government had dealt with the pandemic at wave three as well. But we asked questions such as their self rated general health, the general health question all waves, long standing health conditions, disruption to medical appointments, and difficulties obtaining medication - 46 cohort only in wave one. But as things panned out, we included all the cohorts at wave two, and wave three. And also recorded all waves, whether they had been asked to shield as a vulnerable category as well. Right. So moving on to family and household, just to say, we asked current household composition in house or grid at all waves. And then we asked the questions on things like what relationship satisfaction and conflict in all waves we obviously asked the number and age of children live with. We also asked whether the study member or their partner was pregnant at that wave of the study. And also captured the week of pregnancy at that time as well. And some questions on housing in terms of the number of rooms in their house, postcodes access to gardens and tenure at all waves. in terms of health behaviours. So I didn't really touch on this before, but those included alcohol consumption, physical activity, diet in terms of the quantity of fruit and veg eaten, how they were sleeping as well, because obviously, sleep has become a really hot topic during the pandemic, with people finding sleep patterns disrupted and how they were sleeping and how many hours they on average slept a night as well. And we captured their own. We asked them for their own assessment of their weight at all three cohorts. Wave three, we also added a question on screen time, because we're obviously aware that for some people that they'd been working from home or spending significant portions of their time home for nearly a year. And so it was felt pertinent to ask about the amount of screen time. Then in terms of social contacts, support and loneliness. We asked about contact with friends and family in the last seven days and over a range mediums. So telephone video calls, email, texts, electronic messaging, so that kind of captures how much contact, external contact when in situations when people weren't able to see each other. Also in the frequency that they gave, gave help or received help to anyone outside their household, and participation in online community activities, the sort of street WhatsApp groups that emerged over this time and online quizzing and things like that. And there's as I said the provision of help towards each other, but also whether they felt they had social support as well and also whether they were experiencing loneliness. And then in terms of mental health, we asked the overall life satisfaction, self assessed mental health and control over their life. And then for each of the cohorts we asked mental health and wellbeing scales, and those scales vary by cohort studies, so we use the ones that are routinely asked within that cohort. So for the 46 cohort, and for Next Steps, that's the GHQ varying, whether it's the GHQ 12 or GHQ 28. And then for NCDS, and BCS70 That's the Malaise inventory. And for MCS it's a combination of the Kessler plus WEMWEB plus some social provision questions. And we also ask them just single item question, scales about optimism, risk, patience and trust. And also trusting government and political leaders as we felt that was not just generalised trust, but trusting in the system as well at this point seem to be important. And then as I said, we also ask some life event questions as well. So in terms of the child loop, some of the questions were asked at wave two, and then at wave three was about, primarily about education. So whether the child was enrolled in a school, their school year, and their school type, and the extent of how they managed to do any in person schooling, and home learning. So we asked the hours per day, their online lesson provision, and the extent to which parent, there was parental help with homeschooling for both the self and the partner, and their learning resources. So those were not only those available within the household, but also whether they had access to online provision for extra additional learning resources that they may or may not have had to pay for. And then the parental, the parental assessment of the effect of the

pandemic on their child's academic progress. We also asked the parents to provide their assessment of the child's mental health pre COVID and currently as well. At wave two and three, we ended the questionnaire with asking an open question where it was a free text box answer, in which we asked them the participants to assess for themselves any impact that COVID had had on their lives. And it, we had quite a, particularly in wave three, we had quite a lot of response to that. And then ultimately, at the end, we asked, as I said, at all three ways in order to consent to link to this Zoe symptom, track, app, tracker app, or track apper no tracker app! So all of the three questionnaires are available on on on the CLS website, as Richard said, on our COVID specific area of the website. And I would suggest I've only been able to give you a very brief whistle stop tour of what's, all the rich data that we've collected. But I would recommend that you do have a look at those. Okay, so I think, alright, if I pass it over to questions now, because I think then we're pretty much back on time.

Vanessa Moulton 33:18
Yeah. Hi, Kate. Thanks for that. Um, there are a couple of questions, Christina, if you don't mind the question about response rates, I think Richard might answer that in the next section, so we'll leave that for the moment. And your second one we'll come back to. So one specifically about content. Alison wants to know, I was wondering if parents were asked about their experiences of managing children's behaviour at home?

Kate Smith 33:46
No, No, we didn't. We didn't ask that. No. No, because the most. So the Next Steps cohort are age 30 so they are just embarking on their careers as parents, and then it was mostly the MCS parents who would be dealing with younger siblings of the cohort members. I'm afraid we didn't give space to children's behaviour and the management of that.

Vanessa Moulton 34:21
Okay, and I wanted to know, was data on self harm collected?

Kate Smith 34:27
I think it might have been for MCS, but I am going to have to check that if if not, we have certainly collected data, we have certainly collected data on health harm in the MCS cohorts at age 17 and intend to do so for the next survey. I'm really sorry, I will have to go and check that.

Vanessa Moulton 34:49
And then one more. Were deprivation, Diane wanted to know, were deprivation indices of current addresses considered?

Kate Smith 34:58
Well that would be done via. That could be done via geographical linkage. So I presume that might be a question for Richard, I'm not sure. But I would presume, since we know where the cohort members are living, we would be able to apply that externally it wouldn't be data that we'd collect in the in the questionnaire.

Vanessa Moulton 35:19

Okay, and I'm not sure whether you'll be able to answer this. But do you have the 'n' the number of, it's Zhang who wants to know this. Do you have the 'n' for those completing the child loop?

Kate Smith 35:29
Again, I'm afraid that's not a question for me, it will, it might be worth having a look at the User Guide, although I'm not sure individual level areas. But I mean, we did have we did have very, I mean, if people went through it was a 20 minute survey, or 20-25 minutes survey. So for people, for the relevant people, we don't have any indication that lots of people refused the child loop.

Vanessa Moulton 35:56
And one more on content. Was data on experience of domestic abuse collected?

Kate Smith 36:01
No, it wasn't no, we we made a decision not to collect that, although we know it is fairly pertinent. But it is a very sensitive area. And it's one of those things where, as custodians of the cohorts, and we have to think about the long term implications of any questions that we ask. It's the sort of question that we would, we would address in the main stage of a survey, when we have a lot more contact context, or can offer a lot more support to cohort members, as opposed to one of these sort of like, we would never planned to do this. And so we had to really balance on on on on the balance of risks, we decided not to collect information on domestic abuse. And in response to Diana, I'm not sure it is available, you'd have to explore that with the data team, but it would be possible to do it.

Vanessa Moulton 36:54
Yeah. Okay. I think that's all the questions specifically on content.

Kate Smith 37:00
Okay. That's great. Thank you. I'm sorry, if it was a whistle stop tour, but we sort of like limit slightly limited for time.

Richard Silverwood 37:11
Great, thank you very much indeed for that Kate. Lots of information there. And let me just pull up my next set. Okay, hopefully, you've all got my slides there. So back to me, I'm afraid. And so I'm now going to talk a bit about handling non-response in the COVID-19 surveys. So there's already been a question about this. And we can see that the response levels, particularly in some of the cohorts, and particularly for some of the waves of the data collection, are quite low. And so I think that it means that we need to pay more attention to the extent of non-response and how we handle it in these COVID-19 surveys, then we would do normally, when we're using data from the standard suites of these same longitudinal studies, because in the standard suites, we expect our levels of response and we observe our levels of response to be much higher. So this is certainly more of an issue here in the COVID-19 surveys. That's why I get my own 25 minute slot on this, even though it's only an hour and three quarters, and maybe otherwise, I wouldn't. And so I'm just briefly going to sort of introduce the ideas around this why the non-responses might be an issue, where some potential solutions, think a bit specifically about the target population and the response within the target population, because that's important for how we consider the way in which we deal with non-response. The particular approach

that we have provided within the data sets is non-response weights that can be used in subsequent analyses. So I'm going to go through step by step how we've derived those non-response weights, which might be a bit technical, and isn't necessary for you to know in order to apply them yourselves. But I think it's useful to have some information on what we've done so that you can understand exactly what that is. And quickly just look at one example of how effective they can be in an analysis. And then look at how you will generally use the data sets would implement these non-response weights in practice. So non-response is common in longitudinal surveys. And if you're, you know, you're frequent users of longitudinal surveys, you will know that that is the case in all surveys. And the missing values mean that if you are just to conduct a simple complete case analysis, so only analysing those cohort members for whom you have observed data, then you're only going to be analysing some sub sample of your total sample, a smaller sample size and that means that your analyses are going to be less efficient. But If your respondents differ in some systematic ways from your non respondents, then there's the potential to introduce bias. And it's really this second point, that's the most important. If you have unbiased but less efficient estimates, then I think you will be more prepared to accept that and then the potential for biased estimates. So there's lots of well known methods that have been around for a number of years now or decades, in some cases for dealing with missing data, including multiple imputation, inverse probability weighting and full information maximum likelihood. And I'm a big fan of multiple imputation. So normally, I'd be speaking more about that. But today, I'm going to be thinking about inverse probability weighting. And specifically here when we derive non-response weights, how we can apply those in subsequent analyses. And that's exactly what we've done in the COVID-19 survey. So to correct or to handle non-response, we're providing these non-response weights so that IPW analyses can be undertaken. And the way in which we derive these non-response weights, which I'm going to go into, in a little more detail, really, we're able to do because of all the rich information that we have on our cohort members provided over many years or decades. So because we can characterise our cohort members, whether they've responded or not responded at the COVID-19 surveys, so well, we can capitalise on that very rich information to try and solve this, this issue of non-response. So first of all, a little diversion into thinking about the target population and the response. So the target population generally is really the the population that you wish, your analytical results to be representative of. And what we the way in which we have defined the target population in each of our cohorts for the purpose of thinking about non-response is as individuals who were born in the specified birth periods, because like these cohorts are really defined in terms of the birth period, but those who are alive and still residing in the UK at the point of the COVID-19 surveys being issued. And so the non-response weights that we've derived, are designed to make weighted results from the COVID-19 survey respondents representative of this target population. So that's why it's important to clearly define what the target population is. So we know what we're aiming to make our subsequent analyses representative of. And a slight side note is that the COVID-19 surveys were issued to a relatively small number of cohort members who had already emigrated from the UK. And the main reason for this is because they were often issued by using people's email addresses. And so obviously, if we're just sending information for joining a survey on an email address, we don't really know where people are. But subsequently, we may know that they're not within the UK and therefore did not meet the criteria for our target population it's not not lots of individuals. But we don't derive non-response weights for such individuals. And in fact, we don't use them even in the derivation of the non-response weights for other cohort members. A couple of extra points on MCS, so we've only derived non-response weights for the cohort members and not for the parents. And the reason for this is the sort of definition of the target

population is a bit more difficult for the parents because we don't always know about how many parents there there would be, we don't have information on everyone, and that they're not quite so tightly defined. And we've also only derived non-response weights for singletons and for one twin or triplet from each twin pair or triplets set. So the second twin, and the second and third triplet wouldn't have a non-response weight defined but in fact, subsequent to the derivation of the non response rate all triplet families were excluded from the COVID-19 survey datasets anyway, because there are so few of them and it helps avoid potential issues with kind of identification. So in the first time I spoke, I presented a bit of information on response and there it was relative to the issued sample. So now because we're thinking about non-response, we want to think about response relative to our target population. Our target population here is generally slightly larger than the issued sample because it will include all individuals who may have declined all future participants in the cohort, so wouldn't have been issued the COVID-19 survey. And it will also include some individuals that we've lost contact with. But because we want our results, ultimately to still be representative of such individuals who still remain part of our target population they're included here. We can see that we have similar patterns of response relative to the target population, as we observed relative to the issue sample, so that generally we have higher levels of response in our older cohorts, lower levels of response in our younger cohorts. And as that, as time progressed, within each cohort, the level of response has increased. So how exactly did we derive these non-response rates? So this is an overview of the process that we followed, we conducted this process at each wave separately, and within each cohort separately. So within the sample corresponding to the target population, we modelled COVID-19 survey response, conditional on a common set of covariance using logistic regression. So I will just go into a little bit more detail on this first step, and then come back to the later steps afterwards. So which covariance did we include in this response model, where the covariance that we selected, were informed by results of the Centre for Longitudinal Studies missing data strategy, which is a theme of work that we've had running for some time, alongside all of this, which is helping us to think a bit more clearly and provide clearer user guidance on approaches for dealing with missing data, including the sort of non-response in the CLS cohorts. And that can inform this current work, and also selected based on assumed associations with the probability of response and or associations with key COVID-19 survey variables. So we'll see in a minute a list of the variables that were included. Because we're conducting this process of non-response weight derivation separately in each cohort. We don't have to fully harmonise these variables in the response model across cohorts. But we aim to use broadly the same set of variables in each cohort to ensure consistency. It's not possible to include identical sets of variables, because data collected at different ages, using different questions or measurement instruments. And sometimes the data is just not available at all for one of the cohorts. If you're interested in the full details, including technical details, then they are available in the User Guide. So these are the covariates that we included in the response model. So these are essentially predictors of response in the COVID-19 survey. So there's kind of some basic socio-demographic variables that you might expect, other variables that we've identified from previous work to be predictive of non-response such as whether people will vote, their membership in organisations, whether they give consent for providing biomarkers and linkages to other datasets, a number of variables that we're interested in, particularly looking at, in relation to data from the COVID-19 survey. So it's important to have them in there informing this process around education, economic activity, and a number of kind of health measures. Also, we include the number of non-responses across all previous sweeps of data collection within a cohort, because we've observed and it's clear in the existing literature that this is often very highly predictive of

subsequent response. And then for when we're thinking about response at wave two and wave three of the COVID-19 survey, then we include response at the earlier waves of the COVID-19 survey as a predictor in those response models. So we fit that response model for our COVID-19 surveyed respondents, we predict the probability of response from that logistic regression model, we then calculate the non-response weight as the inverse of the probability of response. We examine the distribution of weights across the cohorts and perform some exploratory analyses to decide whether truncation may be desirable. So when you first calculate these inverse probability weights, the distribution can be such that there's some extreme outlying values that can cause instability in subsequent analyses if you were to use them, so truncation may be desirable. So we've applied a truncation uniformly across the cohorts at a value of 50. So a weight of 50 which is probably fairly liberal, but it unbalanced, it allowed us to still have ways to perform very well in terms of reducing bias due to non response. And we then calibrated the weights or stabilise the weights. So they sum to the number of respondents in each cohort. Okay, um, and so that bit about the process of how we've derived these weights, as I hopefully said, and it's not vital that you, you know, understand or interested in that process. It's for a bit of background. But really the key thing is, you know, if you wish to perform analyses, should you be using these non-response weights? And if so, how would you go about doing that. So briefly now just looked at look at the effectiveness of applying these non response weights, in an example. So we performed a number of sort of test analyses to examine how effective applying these non response weights was in terms of restoring sample representativeness. So a number of these analyses are presented in the User Guide. And here, we'll just look at one of them. So here we consider the distribution of sex in each cohort. So this is observed at baseline in virtually all cohort members. So this variable we already have within each cohort, and it's not really subject to any missingness. What this means is, we kind of know within our cohort sample what the true distribution of sex is. So we can look at this distribution in all cohort members firstly. Then we can look at it in just the respondents to the COVID-19 survey. So this allows us through comparison to our known truth, to examine how much bias is being induced due to non-response, we can then do the same analysis, looking at the distribution of sex, and just within our COVID-19 survey respondents but after having applied our non-response weights, and this allows us by comparing that to the previous set of results, to see how much of that induced bias due to non-response we can remove by applying our non-response rates. Here's the results across each of the cohorts separately. And so in grey, we have the distribution of the percentage female in this case, for all cohort members in red, it's when it's an unweighted analysis of just the COVID-19 wave three survey respondents and in blue, it's the same analysis but having applied the non-response rates. So I think it's perhaps easiest to show in BCS and convenient, the weights performed very well in that cohort. So from the grey bar, we can see that the percentage female in BCS that sort of known true as entertainment was about 49%. If we look at the percentage female, just among our COVID-19 wave three survey respondents, it's about 57%, or 58%. So if we just observe these individuals, and assume that they were representative of all cohort members, we will be someway wide of the mark. And that is implicitly what you will be doing if you just conduct a straightforward analysis of the respondents from the COVID-19 survey without trying to account for potential implications due to non-response in some way. When we apply our non-response weights, we get a percentage female back down almost exactly the same as the known true percentage within that within the cohort. So we can see that in BCS70, applying the weights works very well. If we look at the other cohorts, then we can see that the extent of bias due to non response for this variable, at least it's not necessarily the case, but other variables differs by cohort. So for the older cohorts, when

we're comparing the red bar and the grey bar, there's much less evidence of bias due to non-response than it is in the younger cohort. And of course, of course, because the extent of non-response is higher in the younger cohorts, there's obviously a greater opportunity for that non-response to cause bias. And then by comparing the red and the blue bars, we can see that in each case, the application of the non-response weights is bringing our estimate of the percentage female back towards the known true sample value. But the way in which you performance is a bit variable by cohort, so we can see in NCDS, BCS70 and MCS it works really well. In Next Steps, it still reduces the vast majority of the bias but doesn't work perfectly well. But in other test analyses that we've performed, the weights do work well in Next Steps. So again, see the User Guide if you're interested in more such examples. Okay, so if you wanted to apply these non response weights in practice yourselves, how would you go about doing so? Well, the weights are provided as part of the COVID-19 survey data sets that you can download from the UK Data Service. In cohorts where there are already design weights, so that's NSHD, Next Steps and MCS, the non-response weights have already been combined with those design weights to produce produce a combined weight with that variable name. In other cohorts, there's no, the other two couples, there's no design way but we've used the same variable name for consistency. And so in the way in which you would apply these non-response weights, depends on whether there's other kind of survey design structure to take into account in your analysis. So it depends on the cohort. So in NCDS, BCS70, there's no study designed to take into account. So you just use this combined weight in your analyses. In NSHD there's a design weight, but this has already been combined. So we just use this combined weight index, that's an MCS, there's a bit more structure to account for. So you would begin to use the you would have the design weight, but also want to account for the primary sampling unit, strata and in MCS, the finite population correction. So you just SPY set the data and use the SPY prefix in Stata. So I'm just going to show you how you would do this in Stata but other, you know, would work similarly in other statistical software. So I'm going to illustrate doing this by estimating the proportion of individuals reporting having Coronavirus in each cohort at wave three, well, apart from in NSHD, where it's actually wave two. And so this variable is initially coded using four level categorical variable, sorry and first collapsing them to form a binary Yes, no variable. So the first two categories get collapsed together. And then the final two categories get collapsed together. In NSHD, as I say, this is actually wave two rather than wave three data because even I don't have access to the wave three data at the moment. And so if you're just interested in the proportion of this binary variable, you can use the proportion command and specify this combined weight using the P weight command here, and this if it is just specifying that the cohort is number six, which is NSHD, and we're here we're using a Agresti-Coull 95% confidence intervals, they would often be preferred here, but I don't think we need to get into that at this point. So we can see here that the proportion of cohort members reporting Coronavirus in NSHD at wave two is about 2.2% having waited it. In MCS, the process will be the same. So now using wave three data again, you're just specifying the P weight. In BCS70 as I say, it's the same approach here just specifying a P weight. In Next Steps, you would SPY set the data. So the PSU variable, the strata variable are already so these are the standard ones relating to the initial study design, they're kind of re-provided for you in the COVID-19 survey data sets. So that's why they have this CW3 COVID wave three prefix here, but they're exactly the same variables, as you may be used to using from previous data deposits. So you'd SPY set here making sure that you use this combined weight rather than the standard weight that you would have in Next Steps. And then SPY set prefix for your proportion command. In MCS the same, but there's also this finite population correction to specify here in your SPY set. Okay, so that was pretty much all I had to

say. But I wanted to add in one slide right at the end and just thinking about analysing data across multiple time points, because a lot of research that people will wish to undertake will involve doing exactly this. So it's important to think a bit about how the COVID-19 serving nonresponse ties into this or the considerations that you need to make. So I spend some time thinking about the, about the target population and saying that we design these weights so that we can go from an analysis that just uses COVID-19 survey data amongst respondents, I say respondents at wave three, we can analyse them, we can weight them in an appropriate way so that our conclusions are then representative of our target population. So if we're using data in datasets other than COVID-19 survey wave three, then those same weights, if our ultimate analysis sample, so probably our complete cases across all of our data sources still look at approximately like the COVID-19 wave three survey respondents, then our non-response rates from wave three are still likely to perform well. And this is what we're seeing. So, because the response rates are lower in the COVID-19 surveys, if you're bringing in data from previous sweeps of data collection, when non response was less of an issue, that overall, the complete case sample is largely driven by non-response to the COVID-19 survey. So using those non-response with weights may still be a sensible option. But if your analytical sample doesn't approximately correspond to the respondents, that specific COVID-19 survey, then it's a bit more complicated and alternative approaches may be preferred. So here, that could be following a similar process to what I briefly presented here to derive your own custom weights. But going from your complete case, sample to your target population, possibly using multiple imputation, or full information, maximum likelihood, but I won't go into those details here. Okay, and that's all I had to say. Thank you very much. Are there any questions?

Vanessa Moulton 1:01:27
Okay, I think Christina had one earlier, but, or two actually, but so the first one I think you've answered, but given the low response rates of the youngest cohorts, how representative are they?

Richard Silverwood 1:01:41
Yes, so let me just stop sharing. So I think that, yes, the representativeness of the respondents is obviously, the key concept that I was, that I was talking about here. And so for some of the CLS studies, there is a history of providing non-response weights in this way, and in others that that's less so. And I think that because here, as has correctly been identified, the response rates in the COVID-19 surveys are lower than we would generally see in a normal standard sweep of data collection. And it's more of a concern here. So, yes, if you've only got a response rate of 15, or 20%, relative to your issued sample, then it's certainly right to ask questions about whether or not you know, we think that that we think that they're representative. And, you know, that's that's the whole reason for the work that we've undertaken. So yes, I would say from what we've done, we can see that there are certain of those covariates that we included in our response model, and where there's a clear association between those variables and response to the COVID-19 survey. Given those associations, that means that there are differences between the respondents and the non-respondents in terms of those characteristics. So I think that I would be wary about drawing conclusions, certainly in certain areas, from just the response to the Covid-19 survey, without thinking carefully about how that potentially quite selective response is handled.

Vanessa Moulton 1:03:37

Okay, and then Christina had a second question. Could you please tell us how many cases from Scotland and Wales answered the three waves in the four British surveys?

Richard Silverwood 1:03:46
So no, is the short answer, I don't think I've seen that tabulated. It will be I mean, if you can, if you download the data for the COVID-19 survey waves, which you can very easily do from the UK Data Service when registering on their very simple process and then downloading the data, then you'll be able to have a look, because I think the region is included there as a variable. I suppose it depends, to some extent whether the question relates to country of origin or where they were born at the time of sampling for the cohort study, or whether it's where they are residing at the time of COVID-19. So I think the latter is certainly included, because that was asked about as part of the survey. So I'm afraid I don't have a definitive answer. But it should be easy to observe that for yourself if you download the data.

Vanessa Moulton 1:04:43
Yeah, thank you, Richard. I just had a quick look at the guides to see if there was anything in there but I couldn't I just had a quick look at some data that I've got, but it's just for the BCS and the NCDS and it looks like wave two anyway the country that they were living, Scotland is about 8% percent of the data and Wales is about 5%. So about 500 to 400 Records in Scotland and 300 to 200 in Wales.

Richard Silverwood 1:05:09
Okay, great. Thank you. Okay, any further questions? If we're okay, then? Well, I've only pushed us slightly over time there. So sorry about that. And I'll hand over to Vanessa. Thank you.

Vanessa Moulton 1:05:29
Okay, so um, yeah, thanks, Richard. So we're now going to present two pieces of research in progress using the COVID-19 data in the cohorts. I'm going to start and Bozena will follow with some exciting and hot off the press work on gender pay inequality. We both like to thank our funders, the Health Foundation and the ESRC for this work and the ESRC for Bozena's work and obviously, the cohort members for for taking part in all surveys. So starting with this work, we look at the association between the association between life course trajectories of psychological distress from adolescence to midlife and mental health outcomes during the pandemic. And I'd also like to thank my co-authors. So in terms of the motivation for this project, as as it's been mentioned, a lot. The Covid-19 pandemic and the accompanying policy measures may have potentially affected population mental health. And indeed, some studies in the UK have shown poor mental health in the population, but particularly in the early stages of the pandemic. It's possible this disproportionately affected those who were actually already struggling with their mental health prior to the pandemic. And there is actually some emerging evidence suggests that there has been worsening mental health for these individuals, although it's still open to open to opinion discussion. However, few studies have investigated longitudinal trajectories for mental health prior to the Covid-19 outbreak. Most have employed retrospective measures or prospective measures at one time point or measures over a short period. It has been found that psychological distress in the general general population is shown to be heterogeneous, so it can vary by age, of onset, severity, stability and chronicity. And distinct life course mental health trajectories across across the lifespan, adult lifespan might be related to varying mental health outcomes during the pandemic.

So, as we've been, as we know, one of the main advantage of the of the cohorts is is this wide range of comprehensive data throughout their life, from birth to midlife. And along, we can use this obviously, along with a three waves of the COVID 19 survey. And in this study, or this work we've used the National Child Development Study, born in as Richard said, born in 1958. And who were at the start of the pandemic were around about age 62, and the 1970 British birth cohort study who were aged around 50. And what we wanted to find out were, are differing life course trajectories of psychological distress, but where are they more likely to result in symptoms of depression and anxiety, or lower life satisfactions or indeed, feelings of loneliness during the Covid-19 pandemic. And whether there were differences in these in these outcomes, life trajectories of psychological distress as the pandemic developed. So in terms of the method, so our sample included all participants who were who are not dead or hadn't emigrated by age 50 and 46, respectively, in the two studies, so we have a sample of 15, just over 15,000 in the NCDS, and in the BCS70 17,500. So in each cohort, we conducted 25 imputations and included the main variables of our study and also auxiliary variables. So those that Richard already mentioned, that we use to determine non-response to the Covid surveys, and then others related to the outcomes and the main variable of interest in the study. So our main main variable of interest was pre-pandemic psychological distress at five, five time points. So we're looking at the life course history of psychological distress. And we use latent variable mixture model to identify longitudinal classes. We had three outcome measures across three waves. So psychological distress using the Malaise inventory, which is the same measure that's used pre pandemic, life satisfaction and loneliness using the short version of the UCLA loneliness scale. And we, we use cut off. So we've got binary measures. So for psychological distress, it's four, four or more life satisfaction seven or more and for loneliness, six or more. We also used a wide range of potential confounders in our analysis from early from early life, socio economic, parental child health and child ability. And in terms of looking at the relationship between our pre-pandemic psychological distress and our outcomes, we use more modified Poisson models with robust standard errors, which give us risk ratios. So, our first step was identifying pre-pandemic longitudinal trajectories of psychological distress in our data. And as I mentioned, we used latent class analysis over five sweeps of the cohorts. So from age 16, to 50, in the NCDS, and from age 16, to 46, in the BCS70. And this identifies patterns of psychological distress in the data over a 30 year period. So we ran a series of models on the analytic sample from three to seven classes. And we used a number of FIT (model fit statistics) criteria and prior research to identify the most parsimonious models. So in each cohort, our analysis identified five trajectories of which three were extremely similar. And if you look at at the bottom, the red, the red line, which relates to absence or low symptoms, the top line which is severe repeated symptoms, which is the turquoise, an adult onset with more favourable outcomes, which is the pink dashed line, and then a fourth, which was similar, but different in the two, so the similarity being that they was both midlife onset, and that's the purple line. But with the NCDS, there were more favourable outcomes by age 50. Whereas in the BCS70, you can see that it's increasing, so more severe outcomes. And the fifth, the fifth and final class was the grey line which they were very different in both. So the NCDS are sort of an ongoing minor symptoms and in the BCS70, early adult onset with with more favourable outcomes in later life. So we use these pre-pandemic life course trajectories to look at outcomes during COVID. And this is obviously one example of the type of research you can do with with this with this type of cohort data across the life course. So, just to show you some descriptors for this purposes, here, we're using the mental health outcomes during COVID-19 wave three, the third national lockdown by the by the psychological distress trajectories NCDS is on the top and BCS70 on the bottom and the overall proportions are in the

purple and the purple bar, and from left to right is more or less favourable pretty pandemic life course symptoms. So, for psychological distress overall, nearly 22% of the NCDS and nearly 28% of the BCS70 had high psychological distress at wave three in February March 2021. And this was rising from 5.6 to 59% depending on the trajectory in th NCDS and 8.7 to 70% in the BCS70. Likewise, lower life satisfaction and feelings of loneliness during the pandemic, pandemic varied by life course trajectories of psychological distress. So you can see in both cohorts overall around 30% had feelings of loneliness and this rises from 17 to 55% in the NCDS, and 19% to 60%, and in the BCS70. Interestingly, the majority were satisfied with life so 63% of the NCDS and 58%. In the BCS, we said 70, but this falls from 73% for those with low, low symptoms or absence of symptoms to 43% in the NCDS choose from 70 to 34% in the BCS70. So, we then modelled our variable of interest on our outcomes using, as I mentioned, modified Poisson models, which is a log-binomial link function that returns risk ratios. And the reference category was the lowest symptoms group, so the absence of symptoms. So during wave three, any pre-pandemic experience of psychological distress across the adult life course compared to the absence of symptoms was associated with a greater relative risk of psychological distress trajectories. Not surprisingly, trajectories associated with the greatest risk had more than one prior episode, and more recent occurrences. And, for example, the repeated severe symptoms were associated with a relative risk of 8.6 in the NCDS and 7.4 in the BCS with psychological distress in the pandemic, but you can see there's kind of a varying gradient there in terms of the relative risk. If we look at life satisfaction, so again, all the trajectories compared to absence of symptoms were associated with lower life satisfaction during the pandemic, and the risk in the reduction of high life satisfaction from repeated severe symptoms, for example, was 40% in the NCDS and 50% in the BCS for the midlife onset with severe symptoms the BCS70 likelihood of highlights satisfaction was was actually lowered by a third and again, if we look at loneliness, again, all trajectories compared to the absence were associated with increase relative risk of feelings of loneliness. And for the severe, for example, it was around three times the risk for other groups It was around two or under. So then, here, we compare three time points during the pandemic. And the graph sort of illustrates the differences in the proportion of high psychological distress and each of the groups. And you can see that from each of the trajectories, the proportions of those with high psychological distress were pretty stable over the three waves. But it looks like there's a pattern, perhaps that seems that distress was was worsening over time. And, sorry, scrolling along, okay, um, and if we look at lowness, for example, in the NCDS, there were differences from if you compare wave to wave three. So at wave two it was during, it was when the national lockdown had eased, although there were still social, social constraints, the wave three was during the third national lockdown. So for the NCDS for those with minor symptoms at onset and repeated severe, you can see that there's, there's an increase there in loneliness from wave one, wave two to three. But this difference was found in the younger cohorts. So, to conclude, data from two British birth cohort suggests that the risk of poor mental health outcomes during the pandemic was greater for those with prior symptoms of psychological distress, regardless of the age of onset, severity, and chronicity, particularly for those with pre-pandemic severe symptoms. For most trajectories, the comparisons were, they were distinct relative risks associated with mental health. And the probability of psychological distress and low life satisfaction associated with the pre-pandemic trajectories remained fairly stable at three time points. But as I just mentioned, for loneliness, there was a difference wave two compared to wave three in the NCDS. So I think our findings show the importance of considering heterogeneous mental health trajectories across the life course in the general population, in addition to

just looking at the mental health average population trajectories. So um, that's my work to date. So if there are any questions at all?

Richard Silverwood 1:20:02
I can't see any questions, Vanessa. So may be we should move on. And then if people have any questions, we can address them afterwards in the Q&A session.

Bozena Wielgoszewska 1:20:10
Okay. So, um, so hello, everyone. Thank you for staying until this last presentation. So I'll present a piece of research that's another example that is part of the project that uses the British Cohort Study data to look at the gender wage gap. However, since the COVID pandemic happened, and together with my co-authors are listed here, we're also looking at how this pandemic has affected gender equality. So in terms of the motivation for this research, there's been concerns expressed in the literature that women have been disproportionately adversely affected by the pandemic. And it has gotten a name of a 'she-cession', where previous recessions were called 'man-sessions', because they affected man to a larger extent, and a few possible explanations have been put forward for this. So the first one is that women work in occupations and industries that are more likely to be affected, such as hospitality and tourism. The second one is that the of efficient household production. So this is based on Becker's idea who wrote in 1960s, at the time where women were generally less educated than men. And the idea here is that women have lower earnings potential, and therefore it makes economic sense for them to look after family. And the final idea is that of social norms, which is based on social expectations towards gender roles. So there's debate in the literature currently about the expressed preferences and whether they can reflect people's true preferences, or whether they can be socially constructed. So for example, women being socialised. To think that they are better looking after family as opposed to, for example, being engineers. The contribution that we hope to make with this piece of research is that most studies looked at the effect on gender equality looked at the first lockdown. And we're using for this the wave three of the survey, so looking at it in the medium term, as opposed to the short term. Most studies also focus on couples with children, but we expect that the different mechanisms that can be at play for different family types, so we hope to look at different family types as well. And also at different points of the life course, which are affected by the different cohorts. So our methodology, we use the pooled sample of all four cohorts, we limit our sample to those who are employed at the start of the pandemic, and those who live in England, Scotland and Wales. And we also exclude loans others due to low sample low prevalence in the data. And we use linear weighted weighted linear probability models to do this. And at the moment, we haven't yet got an elaborate missing data strategy. So we're adding a missing category for the missing covariance in order to retain the sample size. So here are our outcomes. So the idea is that when shocks such as COVID hit, who are more likely to be affected in terms of the employment participation, and we're looking at four outcomes so far, the first three are sort of progressively stricter and stricter definition of employment participation. So the first one is anyone who's in employment, so that's employed furloughed apprentice, voluntary work, self employed, as opposed to unemployed, sick, disabled, looking after family, retired or in education. And the second one is more narrow, so it excludes those who are on paid or unpaid leave and those who are furloughed. And here we're only looking at those in active paid work. And the third one are those who are actually doing the same job as they did in March 2020. And the last one is being furloughed. So the way we model this is, we first look at the role gaps, we then add

variables about partner and children, which helps us does this hypothesis of efficient households participation official household hypothesis, and then we add information on both the cohort and some basic controls, the key variables that we know are related to employment participation, and we in the final stage of adjustment as job characteristics, so that's pre-pandemic SOC on one digit level that you can see listed here. And binary variable for whether they work part time or not, that's defined based on their hours worked. We'll look at different types of households, which corresponds to different life stages of the cohort members, as you can see here, so the red ones are NCDS. And they're mostly couples with no children. So that's no children in the household, it doesn't necessarily mean that they haven't got any children, it just means that they don't live with them at the moment, BCS at the light blue, so the largest population, what proportion of those are those partnered with children. And Next Steps are relatively evenly split across the groups, and MCS who are 20 years old now still, mostly alone with no children. Move in, move on to our results. So here are the four outcomes. So the broad and narrow definition of employment doing the same job and probability of being furloughed, and men here represent the black line. And this is female coefficients coefficients on the gender binary variable with the corresponding confidence intervals. So as you can see, effectively, there's no difference in the rates of employment when we look at the broad definition because of the overlap here. However, when we look at those who are active, or those who remain in the same job, we can see that there's gender differences that disappear after controlling for a woman's occupation, or people's occupation. And when we look at the probability of being furloughed, that's the case, women are more likely to be furloughed, even when we take the job characteristics into consideration. So there's no overlap between the yellow line and the black line here, which implies that even though it might not appear on the surface, that there's a difference, when we look to have that sort of deeper dive analysis into it, we can see that this woman's situation in the labour market is a little bit more volatile. And we then look at different family types here. So the reference category are men with children, who are generally the privileged group, the high earners, because they had time to accumulate work experience by the time they have children normally. And also, we know from the literature that they have lower or no child penalty. And we compare these different family types that you can see here to partners, men with children. So starting from the early life, those who have no no children, so this, the last one and third one here, the role estimates are the blue one. So we can see that men and women in this case look quite similar. And the difference goes away when we control for the age. So that's, that's estimates. And then the basic characteristics, the green ones, and the yellow, dope adjustments, don't do that much on top of age. Now for those who are partners, so that's the second one here and the one but last. Here, we can see the job characteristics below estimates make more of a difference for women. But still, it's mostly age, single mothers. As you can see, it's quite a small sample. So quite a large confidence interval. So we can't really conclude much much in in this case, however, women with partnered with partners and children. So that's the first one here that they're less likely to remain in employment, and age and basic adjustments, does very little to change things, but the job characteristics matter. Moving on to the more narrow definition of employment, so those active in paid work when they're alone and have no kids. So that's the rule, blue estimates, they look quite similar men and women. And here again, age accounts for most of the difference, however, still borderline significant for women at this stage. And it's the job that makes women more look like more more like men. And it's similar for partner to those who are partners with no kids. So it's mostly about age for women, and mostly about most of our age for men and mostly about jobs for women, and then the pregnant women with children. Here the difference is a bit more noticeable. And job makes more of a difference. And looking at the same job is a very similar

story here. Women, especially those with children are less likely to remain in the same job unless we take the job characteristics into consideration. Or when we look at the probability of being furloughed things look a bit different. So here, men with no children are the only group that's less likely to be furloughed than men with children as they are on the left on the of this black line. And generally for men, there's no significant difference differences regardless of whether they have partners or kids. However, as a woman, we can see that all sorts of woman in our family types are more likely to be furloughed. And slightly less so the case for those who haven't got kids. And age and basic adjustments here don't change things too much. However, job characteristics mattered the most. And they look borderline like men if they have no kids, but when they have kids, it doesn't matter what job they're doing, they're still more likely to be furloughed. To sum up, following this COVID-19 outbreak, not everybody has been affected to the same extent, and gender and family type makes a difference in this case. So linking back to this possible explanations that they posited at the start, there's some support for the idea that occupations are to blame, because women look more like men when we account for the pre-pandemic characteristics. And these characteristics matter more for women than for men, we can't at this stage really say much about the efficient household hypothesis yet, except that we do see differences by household type. So it is very likely something to do with household. And we plan to do further investigations taking partner jobs into account. However, it is only possible on further subset of observations only on couples, where both partners work. And we see some support for the social norms theory. So this would be the case if we controlled for all all characteristics observable and unobservable, and there was still gender differences. And we can see that the difference, differences are generally larger for men, for women than for men, especially for pregnant women with children. And in the case of probability of being furloughed, we can see that it's more likely for them to be furloughed a woman to be furloughed, regardless of the job characteristics. So let me tell you a little bit more about what we plan to do. And our next step, so this is still quite an early, very early stage of this research, because this is wave three data, which we only just got. And but as I mentioned before, we'd like to look a bit more explore this efficient household helper hypothesis by including partners job and the presence and age of children. Although this is only possible on the sub set of observations, we would like to split the job adjustment also into the occupation component and the part time components to see which aspect is the one that makes a difference for women. And we would like to look at other aspects of working life such as hours worked, pay, and working from home. And obviously, we would like to have a little bit more elaborate strategy when it comes to dealing with missing data, where we hope Richard can advise us a little bit more. So thank you very much for your attention. And here's the link to our website where you can keep up to date with our project and the outputs. We regularly regularly update and my contact details. So if you have any questions or comments, please feel free to contact me. But hopefully I haven't been talking for too long. And there's still some time for questions now. Thank you.

Richard Silverwood 1:33:03
That was great. Thank you very much. And I don't see any questions in the chat specific to that talk. So oh, Alex Bryson has his hand up. Alex, would you like to ask a question?

Alex Bryson 1:33:16

Vanessa? Actually, I was wondering, I don't recall seeing Vanessa, the incidence of your different different life trajectories on mental health. So what percentage were in the really, really bad category, and the really, really nothing ever bad ever happened to them categories?

Vanessa Moulton 1:33:40
Okay, I've got the slide, but it might take a while to...

Alex Bryson 1:33:48
Well, may be tell me afterwards. But it does seem relevant to me, because you spent a lot of time doing these trajectories. And then I saw basically you got five categories. I see people bobbing around in the middle, then I see low down persistently, so people high up persistently, and I'm thinking maybe that really the interest is in the low down and the high up and the bobbing around people are sort of in the middle.

Vanessa Moulton 1:34:16
So I can tell you approximately for both, I mean, the the lower level, around 45 to 50%. I think the higher level are 15% to 20%. And then the rest are in the other three. So Exactly. 50% we're in the middle.

Alex Bryson 1:34:39
Is that true in both birth cohorts?

Vanessa Moulton 1:34:44
Um, they're slightly different. There's more, there's more in the lower level on the NCDS than the BCS.

Alex Bryson 1:34:54
Okay, thank you.

Vanessa Moulton 1:34:56
Okay.

Richard Silverwood 1:35:00
Vanessa, Heather is asking, Is there any evidence on bereavement?

Vanessa Moulton 1:35:07
Right? That's not something that I've looked at at all. So I don't know.

Richard Silverwood 1:35:19
So we have another 10 minutes schedule. So if anyone has any questions at all, then please either raise your hand or put them in the chat. If they're specific to one of the talks this afternoon, we can direct them as appropriate. And we can answer them between us. So any questions at all in relation to the COVID-19 survey data? Or indeed beyond that, if you really want?

Admin 1:35:44

There is a question there for Bozena...

Richard Silverwood 1:35:53
Sorry, there, Emma Wilson, do you have any academic papers on these findings coming out?

Bozena Wielgoszewska 1:36:02
So this piece of research that I presented just now, we're hoping to submit to Longitudinal Studies, special issue on on pandemics. But it's still very early stages at the moment, we hope, we hope that ultimately we will have some outputs. However, there's some work that's already published. So questions from Emma. So yeah, if you'd like to contact me, I can point you to the the papers that have already been published on this, in this area.

Emma 1:36:34
Thank you. Thank you.

Richard Silverwood 1:36:38
And a question from Alex to me. But first of all, Richard Steele is the is the feedback survey available in the chat if people would like to give their feedback on the session this afternoon.

Admin 1:36:50
Yeah, I'm going to put that in now. Thanks.

Richard Silverwood 1:36:54
Excellent. So once that link is available, and if you would like to spend, we'll only take a minute or so just to give some feedback on the session session this afternoon, we'd really appreciate it because it will directly feed into what we do in the future. And so the question to me from Alex was what I would advise on regarding Bozena's requests on imputation strategies. So I think we have already had something of an exchange on this. And so yes, I think that there's several different approaches here. And it's, yeah, it's not a straightforward situation. So here the analysis is pooled rather than cohort specific analyses, I think was that that that was the key thing here. So yes, I think imputation is sensible. Multiple imputation strategy would be sensible. But I think using a pooled cohort analysis, that becomes a bit more tricky, because there's potentially differences between the cohorts in terms of the relationships between the key variables that you're interested in, therefore, you'd want to allow for that. So your imputation strategy might might make more sense being cohort specific, even if the analysis itself isn't. But I think I would need to give that a bit more consideration before giving a full response. But I'm happy to discuss that with you outside of this, certainly. So get in touch. Great, do let us know if there are any other questions. In the meantime, I think we did think of a couple of tricky questions to ask ourselves in the absence of questions from attendees. So one thing that we thought of is that in wave three, as I said early on, responses were by two different modes. So in waves one and two, it was online only in wave three it was initially online with some telephone follow up with cohort members who had not responded, in the online phase. And so I think, then an important question, but someone may or may want to ask would be about how you would deal with those different modes or whether it's important to consider dealing with those two different modes when analysing wave three data. So there's an extensive survey, literature, methodology literature on different modes and mode effects

where the mode effect is taken to mean that you would obtain a different answer to the same question asked of the same individual, purely just because you're asking it by a different mode, and then thinking about how that can impact upon analyses that you undertake. So trying to sort of tease out those mode effects and observational data is very difficult because in this case, cohort members will self select into either online or essentially into telephone because through their non-response to online, so a sort of crude comparison of the item values across those two different modes, between telephone and online is difficult because those two different modes don't contain comparable individuals, they will have different underlying characteristics. So it would, you know, it's very difficult to do you need some sort of control for that selection. Or ideally you would have some sort of randomised experiment with with forced response for that. That's, that's unlikely. So in this, in this case, with the wave three data, what we provided as part of the user guide, the final section on that is just a little bit of information about modes and the potential for mode effects. And what we've provided in the in the appendix is some long tables of just very crude analyses, looking at what we call mode differences. So differences in item values between the two different modes, going through each question question systematically. So these are not necessarily mode effects, because of the selection issues that I have just described. But it may allow us to think a little bit about whether just crude mode differences, how that might impact on any analyses. And so I think that analysing that data, one starting point would just be to adjust for the mode by which the data has been collected in your analyses. And the variable allowing you to do that is available in the wave three data set.

Vanessa Moulton 1:41:29
Yeah, I just like to say that I use the mode in my imputation. And I also used it in in my analysis as a as a binary measure. And it was quite interesting, because for example, it was significant when looking at loneliness, for example. So those who responded by telephone were less, were less likely to feel lonely.

Richard Silverwood 1:41:51
Great. Yeah. So that's definitely a starting point, including a dummy variable for mode in your regressions, as Alex suggests there. But I think, you know, if if there are more nuanced differences between the modes and how they elicit responses from individuals, that may not necessarily be sufficient, but is certainly a good starting point. Any further questions? Okay, perhaps we can wrap up at that point, then? A couple of minutes early. Oh, Alex. Sorry.

Alex Bryson 1:42:37
Yeah, just one quick question. Who else is using the wave three data that we know of? I mean, we now know two people, Vanessa and Bozena. Are they the only ones? Or is there a community of people out there who can mutually support one another?

Richard Silverwood 1:42:57
Um, yes, so there were certainly people within CLS and beyond who are using the data that are and often those are contributing to analyses that are in combination with analysts from other cohort studies. So in particular, some of the things I'm working on as part of the National Core Studies, we are analysing data, including up to wave three in the in the CLS and NSHD data. And combining that with results from other national cohort studies, not all national studies around the country within the NCS. And so yes, there's definitely an analysis going on there. In terms of people external to CLS or UCL

more broadly, I'm not aware of that, again, these data have only been available to download for the last week or two. So I suspect, not that many people externally will be using them yet. But I would hope, and certainly the hope of this session is to provide some information on that and demand and publicise the availability, and the, you know, the potential uses for these. I'm Alex also thinking about the value of a student data set, many will want to use for third year dissertations. And yes, I mean, I don't know to what extent we need a student data set because the the data are very readily available from UKDS under the Standard End User Licence. So yeah, I'm not sure if we need a kind of a subset of that to be available. But yes, I think we would certainly encourage students to consider using these data, in their dissertations and they seem ripe for that sort of analysis. So and I realise I've dominated discussions. there but, does anyone have anything to raise on those issues?

Vanessa Moulton 1:44:58
No, I think I mean, there are a lot of you know, as you mentioned in the introduction, there are lots of guides and, you know, available documentation on the cohorts, you know, COVID that students can use that. Yes, I agree. Totally.

Richard Silverwood 1:45:18
Great. Thank you. So I think we've run out of time there. So I just like to thank everyone for attending today. And hopefully it's been useful, if not, let us know in the feedback form. And we'll try and improve next time. So, thank you all very much.

Vanessa Moulton 1:45:34
Bye. Thank you very much.