UK LLC Data Access Public Review Panel

Friday 17th February 2023

12.30pm – 1.30pm

Attendance	
Kirsteen Campbell	UK LLC Communications and Engagement Officer (Chair)
Robin Flaig	UK LLC Deputy Director (Acting Deputy Chair)
Rebecca Whitehorn	UK LLC Research Administrator
Six Panel Members attended	UK LLC Data Access Public Review Panel
Guest Speakers	
Laura Gimeno	UCL

AGENDA

Agenda Number	Time	Presenter	Agenda Item
1.	**12.30**	**All**	Introduction Updates from previous meeting
2.	**12.40**	**Laura Gimeno**	Presenting project, ref. no: llc_0029 on "Harmonising diagnoses of health conditions across British birth cohorts"
3.	**12.45**	**All**	Questions from Data Access Public Review Panel following llc_0029
4.	**13.40**	**All**	AOB

Minutes

Agenda Number	Presenter	Agenda Item
1.	**All**	**Introductions** **Updates from previous meeting** Application llc_0026 not yet approved. This requires revisions and more input from PPIE group. Applications llc_0027 and llc_0028 are both fully approved.
2.	**Laura Gimeno**	**Presenting application, ref. no: llc_0029 on "Harmonising diagnoses of health conditions across British birth cohorts"** The researcher explained that they are part of the Centre for Longitudinal Studies (CLS) at UCL, home to four of Britain's national birth cohort studies. They are applying to access the UK LLC TRE to produce harmonised data on doctor diagnosed chronic health conditions of people born in 1946, 1958, 1970, early 1990's and early 2000s who are still being followed up. There is value in using these studies as a collection, through comparing generations and observing if patterns in one study hold true in another. These studies began at different times, so data on similar or the same concepts is collected in different ways. In practice, there is a barrier to this work as all studies do not often use one variable to measure the same health condition. This application aims to lower that barrier through retrospective harmonisation. The researcher wants to use data already collected, through surveys and medical records to create indicators, determining if an individual has a particular set of chronic health conditions diagnosed by a doctor. They want to do this in the same way across all studies. Harmonising chronic health condition data is a priority, given the COVID-19 pandemic and the importance of underlying conditions as risk factors for COVID-19 outcomes and Long-covid. The researcher explained that they want to focus on four 'big families' of chronic conditions; cardiovascular diseases, cancers, diabetes and chronic breathing conditions. There are several reasons for focusing on these groups. There is a high chance of being able to successfully harmonise them based on questions asked in the surveys. They are the leading causes of death and disease in UK, so quite common conditions. These groups are also important COVID-19 risk factors, so will be relevant to COVID-19 research and because of their general nature, this will be of use to many researchers.

		Harmonising data and documenting the process properly is time consuming. A lot of research that requires harmonised data may not be done if the data is not available, therefore meaning that research in the public interest is not being produced, and many questions that these birth cohorts could help answer that are not being explored. If every team of researchers carry out their own harmonisation work, they are doing the same work twice but often slightly differently using up time and resources. The researcher explained that by producing harmonised data with high quality documentation, they could accelerate the process of research using these studies, enable more comparative research across the studies and help increase the transparency of the work. This work primarily targets researchers. Because of this and the fact that it will only use data already collected to make methodological improvements, they do not have immediate plans for public involvement. They have access to a PPIE group who work regularly with researchers using data from the 1946 cohort. They hope to be able to meet with them to gain insight on the project. Through this harmonisation work, they aim to make it easier for researchers to use the resources that are Britain's national birth cohort studies for high quality research on health, which can hopefully influence policy and contribute to improving people's lives. While this research will stand to benefit COVID-19 research, it has a broader reach which can help answer questions about broader societal issues such as healthy ageing.
3.	**All**	**Questions from Data Access Public Review Panel following llc_0029** Panel questioned if the process of harmonisation could be easily rolled forward in years to come for future work. The researcher advised that this work is something that is reflected across many different topics in CLS. There has already been work on harmonising BMI, cognition, mental health and education. This project would fit into a larger package and is factored in with new study collection waves. It is quite difficult to harmonise retrospectively, so these studies are now being designed in a way to collect similar data at similar ages. This is a first step, but there are other parts to improve this further. It is important to get a start at harmonisation. Panel questioned why the researcher is looking at Millennium Cohort Study (MCS) when they are younger participants as the application aims to look at chronic health conditions. The researcher advised that the initial objective is to work with these studies as a package. They are aware that it is unlikely there will be large numbers of individuals with these kinds of diagnoses. The harmonisation is in part creating

		these variables, however much of it is also documenting and it is important to establish that baseline. Panel questioned if the fields will be coded data rather than free text. The researcher confirmed. Panel further questioned different coding systems that will be incorporated and asked if they would be 1:1 or 1: many, etc. The researcher advised this is something they would take back to the team. The panel questioned if all the cohorts are of a similar severity, for example, are these population slices rather than a selection of a slice in time. The researcher advised they are born in the same year or within the same couple of years and then followed up through time, so the exposure they share is time of birth. They are not selected on having conditions. Panel questioned how the researcher would deal with questions such as, “when was your condition first diagnosed?” or “how long since your condition was diagnosed?” The researcher advised that this is where bringing in linked health data will help. The first stage is identifying as of now, has this individual ever had a diagnosis within this set of conditions, and then will work back. Through the process of documentation, this should become clearer. The panel questioned biomarker data and their use of blood pressure as an example. They questioned what use the researcher would make of these biomarkers. The researcher advised that they had initially considered bringing in many types of data such as procedural and prescription. They chose to strip this back to doctor diagnosis data in the first stage, as this is closest to what has been asked by studies. The biomarkers are a later validation comparison documentation aspect. Panel asked if the researcher has spoken to any researchers to see if harmonised data would be of use to them. The researcher advised they have not yet but would find value in this. They further advised that the CLS have recently had three covid sweeps across all five studies, asking questions in similar ways. The panel further asked if they have any documentation to show they have engaged with these people. The researcher advised they do not at the moment, but they can try to collate evidence. Panel asked if this would be re-run on refreshes of the data. The researcher advised this is something they are discussing with the data management team at the CLS. They hope to have a method in place, so it is easier to run updates, and this is something that is maintained. Panel asked if harmonising data is limitation, and would they lose the unique meaning of the original variables between different cohorts.

		The researcher advised that this is a major issue with harmonisation as it is balancing between keeping enough detail and having enough points of comparison. They see harmonised variables as being created and made accessible to researchers, but this would not invalidate or take away the ability to look at the original ways of coding. This is enabled through being transparent and documenting the process. The panel asked for an example of how different variables from a condition such as cancer will be harmonised. The researcher explained that the first point would be lots of documentation, meaning going through all questionnaires in all cohorts. This includes writing out the full questions, thinking about the order in which the questions were asked, what the possible response categories are. They would then collate these, identify the topic and then the first step would be to look at these questions and harmonise from that. They would then bring in codes from linked data using these code lists and then there would later be the validation process. Panel questioned if the 1946 cohort PPIE group is made up of a diverse group of individuals representative of the general population. The researcher advised that they do not think it is and that it is reflective of that cohort. They advised they are interested to hear any suggestions that people may have around this. Panel asked if they could see variations in COVID-19 complications and outcomes based on the lifestyles of different generations. For example, smoking in the 1940s and 1950s compared to today, how does this reflect diagnoses and health data usage now. The researcher advised that they will not be looking at anything directly COVID-19 related, they are instead trying to identify chronic conditions. Panel asked if the researcher would be looking at correlations between variables such as gender and ethnicity. The researcher advised that the idea is to build harmonised variables and before releasing to the public, they would ensure the quality is good enough for others to use. An important part of this is a description of how these differ through the cohorts including by gender. This would be encompassed in the validation. Panel noted that the lay summary does not reflect the aims, objectives and priorities of this study. It mentioned COVID-19 outcomes; however it was mentioned in the presentation that this is not of particular interest. The researcher advised that this is something that can be adjusted so it is clearer.
Feedback and outcome		• **Revise lay summary to reflect and clarify COVID-19 relevance**

		• **Consider speaking with other researchers about this work and the benefits to their research – keep documentation of this process** • **Consider further PPIE groups, other than just the 1946 birth cohort**
8.	**All**	**AOB**