UK Health Data Research Alliance – Improving Access to Linked Data for Research

Workshop Summary – 27th April 2023

On 27th April 2023, the UK Health Data Research Alliance (the Alliance) convened the third and final event on 'Improving Access to Linked Data Research', chaired by Paola Quattroni (HDR UK, Head of Alliance Strategy and Engagement). This workshop followed two previous sessions focussing on challenges and opportunities around using linked data for research, summarised on the Alliance website.

This last workshop delved deeper into the themes that were previously highlighted, aiming to gain a better understanding of data linkage, the use of consented and routinely collected data, and the development of cohorts/models for longitudinal studies.

More than 60 participants from the health-related sector attended the meeting, including research groups, regulators, custodians, and clinicians.

The session aimed to achieve the following objectives.

1. Identify opportunities for improvements to enable access to and use of consented data linked with routinely collected health data.
2. Establish the start of an opinion piece addressing key issues to overcome, whilst identifying volunteer contributors.

All presentations from the 'Improving Access to Linked Data for Research' workshop can be found in 'outputs' on the Alliance website.

Why is linkage important when collecting clinical cohorts?

The first session was led by Reecha Sofat, Head of Clinical Pharmacology and Therapeutics, University of Liverpool / Associate Director, BHF Data Science Centre).

Reecha's slides and presentation focused on her experiences as a researcher and why data linkage is important when collecting disease base clinical cohorts and the different ways methods for understanding causes and consequences of disease (existing primary and linked health records; population cohorts with incident disease; randomised controlled trials' case-control studies; prognostic cohorts).

Integrating contented and non-consented data for longitudinal population studies

The second session was led by Andy Boyd, Director of the UK Longitudinal Linkage Collaboration (UK LLC), who provided the perspective of a data custodian, exploring 'Integrating consented and non-consented data for longitudinal population studies'.

Andy's slides and presentation followed on from his presentation at a previous workshop and focused on longitudinal studies and 'consent' (in the context of longitudinal research) as well as highlighted the work undertaken by UKLLC.

Data linkage with consent

The second session was led by colleagues from Our Future Health (Kate Evans, Jackie Shears & Fiona Maleady-Crowe) and explored 'data linkage with consent'.

Kate, Jackie and Fiona's slides and presentation focused on the objectives of Our Future Health (to become the UK's largest ever health research programme, with over 5 million volunteers sharing information) and the challenges and opportunities with data linkage and how the three pillars of consent underpin all of this work.

Main discussion points

Questions were welcomed from contributors to discuss the presentations, guiding the conversations summarised below.

Importance of linkages

- The participants engaged in a discussion regarding the granularity of linking factors for socio-demographic data, including air pollution. They specifically addressed the following questions:

1. What linking factors are used to connect socio-demographic data, including air pollution?
2. Is the GP Postcode employed for linking purposes?
3. Is there a household key or identifier used in the data linkage process?

- Linking factors vary greatly as each study has different trust relationships based on commitments to patients and the research question. Where possible, participants are linked up via their property addresses and the allocation and exposure at that level. Some studies prefer this to be done via postcode level.
- There is a large HDR UK programme called 'Social and Environmental Determinants of Health' which is looking broadly at all health data and cross-cutting health data science. It will consider the unique property number and how this can be used as a universal 'household key' across data science platforms whilst retaining privacy controls. This is being used in SAIL Databank and will hopefully be used more widely going forward.

Importance of harmonising data sources

- During the discussion, the question arose regarding the use of the OMOP common data model to standardise and harmonise the various data sources.
- The participants inquired whether this model was employed for achieving consistency and coherence. If not, they sought to explore alternative approaches used for standardisation.
- Our Future Health are considering the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) and the NHS Digital (now England)'s cross-mapping coding system. This system has not incorporated the common data model. There is a much wider conversation as to how the UK health system react and bring these different data models together in harmony. They need to serve the researchers and the patients in the best way possible.
- UKLLC are also considering how this will best work. Born in Bradford have implemented the OMOP common data model and have had a positive experience. There are many strong examples of where this works well in the UK and across the world.
- There are legitimate questions about whether OMOP is the correct model for longitudinal data. The researcher's experience needs to be improved and the way of doing this is by better understanding the different data models and standardising where possible.

- The Alliance will bring together the community to further explore the OMOP model and others to consider the correct solutions with a Special Interest Group on OMOP CDM.

Improving participation and public trust

- It was acknowledged that bridging the gap between participants and researchers is necessary. This is a task of huge difficulty. Our Future Health are working closely with participants to develop plain language summaries of research proposal that clearly explain the public health benefit. Researchers are trying to demonstrate how use of data for which participant consent has been obtained delivers patient benefit. It is important that researchers and participants communicate and build on these efforts further.
- During the discussion, the topic of participants regretting their decision of consenting for use of their data arose. Specifically, the participants explored whether there were any concrete instances where individuals expressed regret after participating in the study, possibly due to gaining 'too much' knowledge about their future health. There is a significant piece of work being undertaken which involved substantial engagement with professionals and professionals to explore how best to share difficult and surprising information with participants.
- There are examples on how this can be done well, including Genomics England. The 100,000 genomes project included an optional consent section that allowed the researchers to look beyond the initial areas of interest. Before these findings were shared with the participants, they were given the option to not receive it. Our Future Health are learning from these examples and exploring more ways of doing this in varying circumstances. There is also risk of participants of misunderstanding the findings. It's vital that the information is very clear.
- Participants explored effective strategies for communicating with patients or participants about the potential future data linkage once a research project has been completed, particularly in situations where there is no active communication channel available. Within the cohort community, there are many participants that have deceased or don't have the capacity to fully understand the full extent of the research. There are no concrete mechanisms to best manage the ongoing duty of informed consent and there is a great need to improve the existing practices. Historical cohorts in particular have huge difficulties in reconsenting their participants to use data for future studies. Prospectively, when starting new cohorts, this should be built into the consent.

Next Steps:

HDR UK and Alliance members will be working towards an opinion piece that addresses the topics discussed in this series of workshops and will highlight areas where further improvements might be needed. The aim will be to encourage relevant communities within the health data research ecosystem to work together to overcome the barriers in linking consented data with routinely collected health data.

Those that wish to contribute to this paper are asked to contact us via ukalliance@hdruk.ac.uk **no later than 2nd July 2023.**