

Workshop Summary

Background

The value of effectively and safely linking consented data (that requiring active assent from those providing it, e.g. the donation of biological specimens) with routinely collected medical data for research was highlighted by the UK Health Data Research Alliance (The Alliance). It was acknowledged that researchers face many challenges when attempting to achieve this in practice, including inconsistent application procedures, long timelines for accessing datasets and difficulties articulating the benefits of using the data and meeting requirements for ensuring secure working practices. These challenges may result in missed opportunities for investigating key health and social care research questions and, by extension, for improving treatment options and the quality of care patients receive.

Some of these challenges were experienced by many Alliance members, prompting the formation of a working group to bring together a broad tent of actors within the health data ecosystem to strategise around making headway in this complex and sensitive issue.

The linking data for research working group

On 13th July 2022, the Alliance convened a workshop chaired by Meredith Leston (Nuffield Department of Primary Care Health Sciences, University of Oxford) to work towards the following objectives:

1. Identify opportunities for unlocking the value of broad linkages between consented and routinely collected health data for public health and discuss key roadblocks to achieving them
2. Explore collaborations under this remit across Alliance members and partner organisations
3. Develop recommendations to drive improvements in the linking of consented and routinely collected health data

Over 50 representatives from health-related sectors, research groups, regulators and interest groups attended this meeting and listened to presentations provided by Dr Nick Thomas from the Royal College of General Practitioners' Research and Surveillance Centre, David Jenkinson from The Brain Tumour Charity, Sneha Anand and Meredith Leston from the Oxford-Royal College of General Practitioners' Clinical Informatics Digital Hub (ORCHID) and Andy Boyd from the UK Longitudinal Linkage Collaborative. Collectively, these presentations summarised the key use-cases, opportunities, and challenges that health researchers, data custodians, clinicians and advocates experience when attempting to form or work with these types of health data linkages or access their constituent datasets.

Below we outline the main takeaways of this meeting and the insights and recommendations for change that emerged:

Clinician Perspective: 'Sharing and Linking Health Data from the Clinician's Perspective: Pitfalls and Myth-busting'

Nick Thomas, Clinical Research Lead at the RCGP RSC and General Practitioner at Windrush Medical Practice, highlighted the health data linkage challenges that are most prominent in primary care settings:

- Many different types of organisations are currently asking for data access without specifying exactly what data they, why they need it and how it's going to be used and stored. This creates major trust issues for both patients and clinicians. Bolstering trust must be a priority and must be incorporated into data access requests.
- More outreach must be done to reassure practices that they are legally protected when safely providing access to data and to publicise the success stories that come out of this work for clinicians and patients alike.
- The language used to describe data sharing needs to become more accessible; plain English summaries of data sharing agreements must be provided alongside legal documents and terminology such as 'Trusted Research Environments' must be demystified.
- There is room for improvement in data sharing between primary care and secondary care settings.
- Many patients have health data records across various sites and settings. It is often difficult for health professionals to access patient health data that has been collected in a different setting. There are lengthy processes in place to allow this data sharing, but it often delays patient care and is even more so for research.

Charity and Patient Perspective: 'Just use my data'

David Jenkinson, Chief Scientific Officer, The Brain Tumour Charity highlighted the importance of timely health data linkage for brain tumour research. In this moving presentation, he shared research The Brain Tumour Charity has completed on patients' motivation to share their data more widely to unlock the treatments their lives depend upon. Here:

- 97% of brain tumour patients said '*I would be happy to give my medical health data to help improve brain tumour treatment and care*'.
- 94% of brain tumour patients said '*I would be happy to share my information even if I could potentially be identified from it*'.

Although all patient data must be handled with due care, this presentation served to emphasise how some patients unfortunately cannot wait for researchers to go through extensive bureaucratic processes to access and link data and finally begin pursuing new treatments. Timeliness must therefore become a key metric of effective data access and academics must be

made aware of the availability of these datasets and the urgency of their usage. Regular communications are needed to reassure participants that work is being done to harness the information they have provided for their benefit and that of patients just like them, present and future.

Researcher Perspective: ‘Opportunities and Challenges Associated with Diverse Health Data Linkages: Perspectives of those Working within a Trusted Research Environment’

Sneha Anand and Meredith Leston from the ORCHID Trusted Research Environment within the University of Oxford provided insight on the issues that academic health researchers face when trying to complete research projects that leverage and link diverse health datasets. They recommended the following steps be taken:

- Regardless of setting, clinical data must, where possible, be codified consistently – this should be incentivised at all levels of health care and superfluous codes should be retired and clinicians made aware when this occurs.
- A central repository of code types should be made available to make it clear how various codes map onto each other e.g. how influenza-like illness (ILI) is coded in SNOMED vs READ.
- The loosening of linkage restrictions that occurred during the COVID-19 pandemic should remain in place.
- Data access request requirements for researchers must be made consistent across data types or, where possible, use-cases (e.g. same timelines/ documents needing to be filled).
- Those making data access requests must be absolutely clear on what they need from data custodians/ those responsible for curating data and need to be able to articulate the public benefit of using that data.
- Work must be done to identify the datasets that, when triangulated, inadvertently reidentify patients; a tool should then be designed to flag when researchers are at risk of doing this.
- There needs to be health data consensus between England, Scotland, Northern Ireland and Wales.
- Terms of consent for data use and share must be defined – how broad should consent be? How far should consent be stretched beyond point of consent?
- GDPR must be more accommodating for sharing international health data.

Data Custodian Perspective: ‘Integrating consented and non-consented data for longitudinal population studies’

Andy Boyd, Director of the UK Longitudinal Linkage Collaborative (UK LLC), provided us with a data custodian perspective on the intersection between consent and forging diverse data linkages when executing longitudinal research:

- Longitudinal studies look at groups of individuals or factors and follow them up over time, regularly collecting data on them. This information allows researchers to

investigate the relationship between different things that occur in individuals' lives and how this can lead to changes in their health and wellbeing or personal circumstances.

- UK LLC makes clear to study participants that they never use personal identifiers such as their name or address in any research; instead, proxies such as NHS numbers are used to link study information to health records. These diverse data linkages enable study information to be contextualised by a participant's entire health profile and onward health outcomes.
- Participants must be made aware that none of this work is conducted for profit-making purposes. Data is not sold and never will be; it is managed internally and is never outsourced to private companies. Doing this cultivates the trust that is the bedrock of a viable data linkage ecosystem.
- However, consent is fragile and can be inadvertently biased if it excludes those too vulnerable to give informed consent from benefitting from research powered by data linkages.
- Opt-in, opt-out and meshed consent offerings should be pursued, and forward planning is needed to ensure that consent is robust across participant life stages (e.g. what additional measures are needed to alter consent taken from a participant's parent at birth to make it viable as said participant becomes an adult and then dies?). All these means of obtaining or stretching consent must be co-developed with participants themselves.

Possible priorities for this group

The Q&A and discussion at this meeting highlighted three key areas of focus for future workshops.

1. It is vital to communicate the value of using linked data for research to the public in an accessible way; success stories should be widely promoted and security arrangements including Trusted Research Environments must be demystified.
2. The complexities that surround the data sharing/access agreements required to execute this work can generate suspicion amongst the clinicians who guard routine medical data and create undue pressures in their already busy working lives. This means these documents often never get signed or executed. There is therefore a clear need to streamline and simplify these legal processes and provide wraparound support or guiding documents to navigate this process.
3. More must be done to understand the unique complexities that surround different types of consent; consent itself may not be the appropriate framework for executing these linkages given issues around how broadly initial consent can be stretched, whether informed consent has been obtained and when ensuring those who can't give informed consent still benefit from linkage work. Currently, different datasets adhere to different burdens of proof in terms of consent; consistency must be achieved to improve the interoperability of health data.

Following this first introductory meeting, we will hold further working sessions to explore these three key areas in more detail and consolidate the benefits of wider health data linkage that were raised and their current barriers into a list of recommendations for relevant decision makers to consider.

The next workshop on 'Using linked data for research' will take place in October 2022. The exact date is to be decided imminently and invites will be sent out in due course.