

Unlocking linked real-world data presents opportunities to improve public health

The COVID-19 pandemic has surfaced the potential and risks of linked real world datasets to accelerate and produce new improvements in public health. In this post, [Matthew Franklin](#), [Dan Howdon](#), [Suzanne Mason](#), [Tony Stone](#), [Monica Jones](#), outline the opportunities and challenges of using real world data as part of the '[Unlocking data to inform public health policy and practice](#)' project. Highlighting the ethical and practical challenges of accessing this data, they argue investing in and developing trust across those involved in the formation of real world data is critical to its effective use.

The use of the terms “real-world data” (RWD) and “real-world evidence” (RWE) in the context of health decision-making has [grown substantially](#) in the last 20 years, although unified and consistent definitions of these terms remain elusive. Often referred to as ‘administrative’, ‘observational’, ‘routine’, ‘large’, or even ‘big’ data sources; over the last decade they have become of increasing interest to those conducting [health technology appraisal \(HTA\) processes](#) to provide policymakers with evidence to inform decision-making and develop guidance on the reimbursement and administration of new health technologies within a care system.

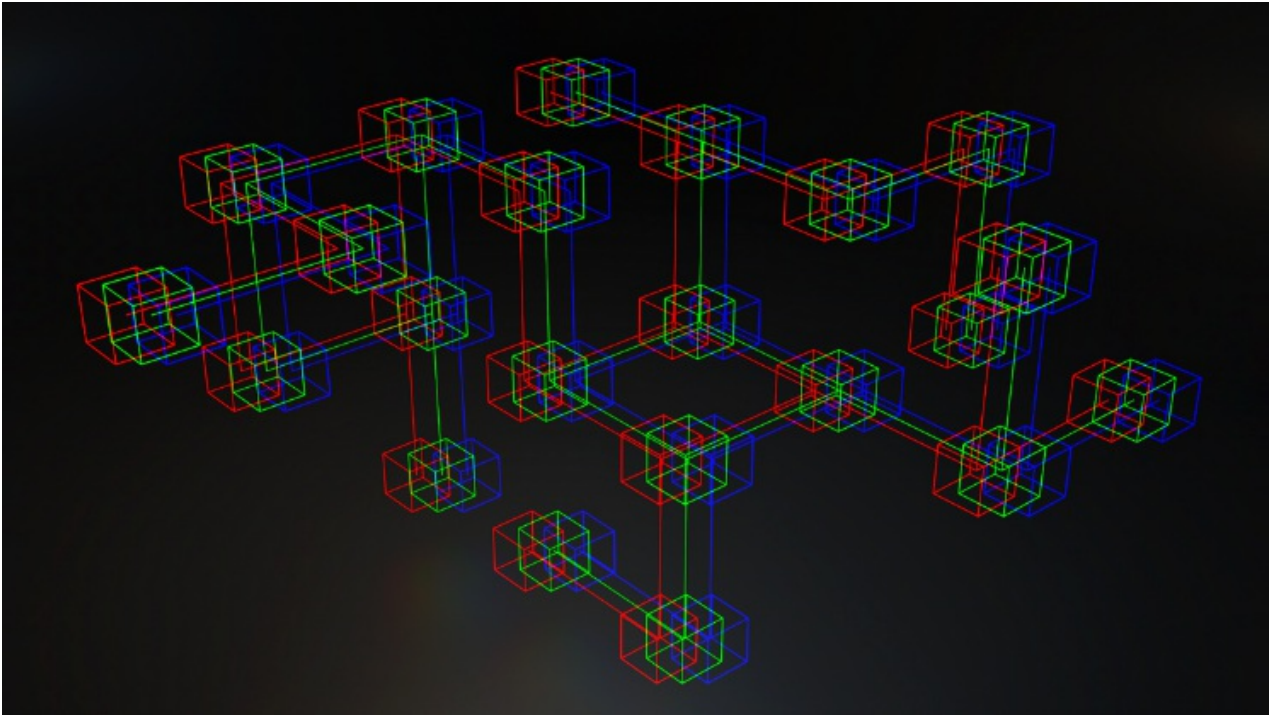
In general real world data and evidence are now used as terms to encompass data and evidence emerging from [non-interventional sources](#), or from sources [other than Randomised Controlled Trials](#) (RCTs). This includes administrative data (e.g. [Hospital Episode Statistics](#)), or survey data on populations (e.g. [Health Survey for England](#)), which can comprise of standalone or linked datasets. Compared to the often perceived ‘gold standard’ of RCT data, real world data presents particular challenges, especially around data protection. However, it can complement or stand in for RCT data. Analysed in its own right, it can also provide descriptive information and be used to assess perceived associations between factors (e.g. to what extent a person’s or groups frailty status is associated with their quality of life, and care resources required and consumed).

Compared to the often perceived ‘gold standard’ of RCT data, real world data presents particular challenges, especially around data protection

As part of an NIHR funded [Unlocking Data project](#), we have been exploring sources of such RWD, often held by local agencies such as councils and clinical commissioning groups in England, and how enabling broader and transparent use of this data (e.g. for research purposes) can be used to promote and protect health, and prevent ill-health.

How real-world data can be used to promote and protect health and prevent ill-health

As stated in the [Life Sciences Vision](#), [Life Sciences Industrial Strategy](#), and [NIHR Best Research for Best Health](#), unlocking the potential of RWD provides huge opportunities to understand and provide solutions for improving health outcomes of patients and populations, informing the development of interventions that optimise disease management and treatment. Addressing these challenges requires partnership between the data collectors (e.g. NHS diagnostic labs, NHS data providers, social care, registries, private providers), owners and guardians of the data (e.g. health and social care providers, commissioners, NHS Digital), and health data users, including researchers, patients, and service providers. Working together to identify and overcome the barriers that exist in accessing data from multiple sources for secondary use will be key to successfully developing linked datasets at scale for secondary use.



RWD can be linked to create datasets that address population health by reflecting the whole spectrum of care experienced by patients regardless of organisational boundaries. For example, we can link data for palliative patients from sources such as hospices, with secondary care data to identify where patients might ‘fall through the gaps’ in care – and then establish how such gaps can be plugged leading to improved quality of life.

Linked datasets are also useful to identify at-risk cohorts, describe inequalities in access to care, model different possible outcomes to treatment, deliver efficient trials (e.g. for understanding compliance with drugs, drug interactions and repurposing of drugs for other uses), apply advanced statistical techniques (such as machine learning) for better risk prediction, or developing clinical decision tools.

Challenges to making better use of real-world data

Information systems are designed to efficiently deliver a specific service. Less consideration is given to integration with other systems leading to fragmentation of data within and between organisations. Documentation (metadata) of source systems, their functions, data stores and flows is crucial to understanding what RWD exists and how it can be used.

The [UK's National Statistician has written](#) that, “Being able to link data will be vital for enhancing our understanding of society, driving policy change for greater public good and minimising respondent burden.” The [UK Government](#), the [Office for National Statistics](#), [ADR UK](#), and [HDR UK](#) all have corporate strategies that include increasing use of linked RWD. This will require the sharing of data across organisational boundaries.

UK data protection legislation does not forbid considered and proportionate sharing of personal data for limited, clearly justified purposes. However, protections exist under common law for information provided in confidence (e.g. disclosed to a doctor in the course of a consultation; or provided to a local authority in connection with their functions). Without obtaining consent – which may be impractical or impossible – such information cannot be shared without risk to the sharing organisation(s) unless through a specific legislated gateway.

Identifying, agreeing and documenting data sharing initiatives is not routine practice. In case of doubt, organisations are likely to avoid the additional risk of sharing data but also miss the [potential benefits](#). As such, there is also the need to build [trust](#) in the use, linkage, and sharing of data

How can we responsibly unlock real-world data

The COVID-19 outbreak has heightened the need for regional, national and international population health management; this has led to significant developments in the use of RWD e.g. the regional development of the Yorkshire and Humber Care Record ([YHCR](#)). Using real time, real world evidence we have a better chance of tackling the challenges posed by a global pandemic. For example, making data more discoverable has been a key aim of HDRUK through the [Innovation Gateway](#) alongside developments, improvements, and access to [care metadata](#). From this then access, analysis, dissemination and transparency for research becomes possible, representing an important aspect of infrastructure which continuously needs 'levelling up'.

Using real time, real world evidence we have a better chance of tackling the challenges posed by a global pandemic.

The NHS has recently published, in draft form, a single Data Strategy for health and care: '[Data Saves Lives: Reshaping Health and Social Care with Data](#)'. This together with the Life Sciences Vision and Clinical Research Implementation Plan, envisages much more widespread use of data the health and care system generates day-to-day in driving insight to support population health, resource planning, clinical research and health-improving innovations.

To enable this the role of [Trusted Research Environments \(TREs\)](#) in the health and care system is being defined. TREs are controlled digital environments used to store and analyse sensitive data securely. The main benefits are improvements in data quality, security, transparency and privacy. Data only resides on systems owned by accredited partners and every interaction is recorded and audited.

In addition, conformance to consistent standards on access and governance is important so that patients, the public and health and care professionals can understand what they do and how they work, and have confidence that they control access to data securely. Through this approach we can unlock the real-world data for the benefit of all. In the short term, collaborations (e.g. between researchers in the NHS and universities) on further data analysis projects are required to fully understand the impact of the global pandemic. In the longer term, collaborations will broaden, relevant skills sets across disciplines will grow and merge, and the use of TREs will allow analysis on a range of healthcare questions to improve patient outcomes and save lives.

Readers interested in linked data and its related challenges may also be interested in these two video animations focused on, the [Benefits and Risks of Patient Data Sharing](#) and [What Happens to My Patient Data?](#)

The '[Unlocking data to inform public health policy and practice](#)' project was funded by [NIHR Public Health Research \(PHR\)](#) with in kind support by the [NIHR Applied Research Collaboration Yorkshire and Humber \(ARC-YH\)](#).

Note: This article gives the views of the authors, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our [Comments Policy](#) if you have any concerns on posting a comment below.

Image Credit: Adapted from [Shubham Dhage](#) via Unsplash.
