# Linking and sharing routine health data for research in England

**June 2017**

Dr Charlotte Warren-Gash
charlotte.warren-gash1@lshtm.ac.uk

Research using de-identified routine health data can provide unique insights to improve population health. England has a wealth of routine health data from sources such as electronic records and public health surveillance systems. However, due to challenges in linking and sharing these datasets their potential to enable powerful, efficient research that informs health policy and services is not being realised. How can we gain the greatest benefits to society from the collection of routine health data?

## Key facts

- Healthcare delivery in settings such as hospitals and GP practices generates large amounts of routine health data, which have important secondary uses for research

- Routine health datasets are used to monitor trends in diseases over time and between countries, regions or healthcare settings, to identify individuals at risk of particular diseases and to inform health intervention strategies

- Research using routine health data includes information on groups typically under-represented in research studies, such as older or vulnerable people. Using linked routine datasets can improve the accuracy, completeness and power of research studies

- Controls and safeguards placed on certain data for the legitimate purpose of protecting the identity of patients, can create unintended barriers to other appropriate uses such as research

- Challenges impeding the use of de-identified extracts of routine health data for research include organisational culture, lack of transparency and communication, complex governance arrangements and technical barriers

## What are routine health data?

Healthcare generates large amounts of routine data for clinical and administrative purposes in settings such as hospitals, laboratories, GP practices and pharmacies.

The storage and flows of routine health data are complex. Electronic health records capture data on diagnoses, investigations, treatments and referrals for clinical management. Public health surveillance systems help to monitor outbreaks of infectious diseases, protect population health and inform service planning. Other routine data sources include birth and death registrations, disease registries and national screening programmes.

As routine health data are not primarily collected for research, linkage between several different data sources might be needed in order to improve data power and quality to answer particular health research questions.

## How and why are routine health datasets used for research?

Routine health datasets are used to monitor trends in diseases over time and between countries, regions or healthcare settings and to inform health intervention strategies.

In the area of infectious diseases, research using routine health datasets might involve monitoring the number of cases of HIV and TB, and investigating risk factors for acquiring infections or complications after infection. Other uses include evaluating the coverage and effectiveness of screening, vaccination or treatment programmes, and monitoring the use of healthcare services associated with infections.

Advantages to using de-identified routine health datasets for research include their large size and real world nature. Information is captured about groups who are typically under-represented in research studies, such as older people with multiple medical problems and vulnerable groups such as migrants or homeless people. The ability to link between datasets further improves the accuracy and completeness of data available for research. Also, using existing data minimises the cost and logistical challenges of data collection.

## Challenges to using routine health data for research



Organisational culture

Lack of transparency and communication

Complex governance arrangements

Technical barriers
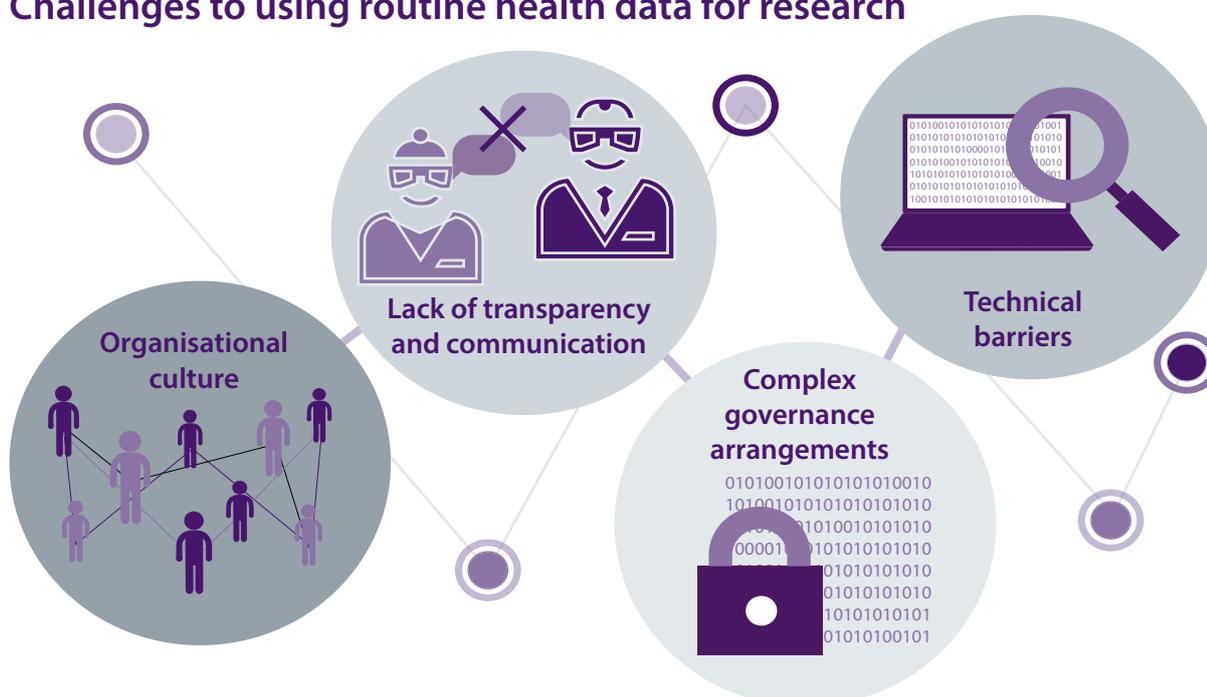
## What are the challenges to accessing and using routine health data for research?

Semi-structured interviews with data users identified the following challenges:

### Organisational culture

Some organisations that collect and collate routine health data do not prioritise external research and may lack the capacity to support it, especially if research is not a core organisational function. There may also be concerns about the legitimacy of data sharing, with issues of competition, control and desire for reciprocity further hampering research.

### Lack of transparency and communication

Data access models vary across datasets and organisations. In some cases, no information is available to external researchers about the datasets collected or the costs and processes needed to access them. There may be no clear point of contact and data access may rely upon personal relationships.

### Complex governance arrangements

Appropriate legal, ethical and governance frameworks are essential to protect routine health data. However, risk averse practices in some organisations may result in complex sets of permissions needed to access routine data. As more datasets are linked together, access procedures become more complicated and time-consuming and may be disproportionate to the risks involved.

### Technical barriers

Substantial work is needed to clean, transform and manipulate routine health data before analysis can begin. Analysing and interpreting routine health data is not straightforward and requires good subject, epidemiological and technical expertise. Interpretation of research results may be hindered by opaque data linkage processes. Lack of sharing of research methods and tools can result in duplicated efforts.

### Data sharing case study - Scotland

There are fewer barriers to accessing national Scottish routine datasets for research. While this is helped by having a smaller population size and longer-term stability and investment in data infrastructure, other features of the Scottish system that facilitate data linkage and sharing for research are:

**The electronic Data Research and Innovation Service (eDRIS),** which offers tailored support for researchers using routine health data. This includes a research co-ordinator who assists with study design, approvals and data access.

**A user-friendly website describing the national datasets available**. Further detailed information about the variables included is available on request.

**A streamlined approvals process** through the Public Benefit and Privacy Panel (PBPP), which scrutinises all research using NHS-derived data in Scotland.

**Use of the Community Health Index (CHI) number as an identifier** across multiple routine datasets which reduces technical barriers to data linkage.

## Maximising the benefit to society

Ultimately, we should aim to maximise societal benefits gained from the collection of routine health data in England. These include facilitating the conduct of research using routine data to **inform improvements to health and healthcare systems**.

Enabling such research involves creating structures for secure data linkage and sharing that operate within clear **legal and information governance frameworks**, while being transparent and responsive to user needs. As the volume of routine health data continues to grow, it is imperative that barriers to access and use do not limit its utility for public benefit.

---

### Key recommendations

Policymakers within governments and public organisations concerned with health data should:

- Establish systems and incentives to encourage secure data linkage and sharing for research in England

- Increase capacity for data linkage and sharing by public organisations

- Streamline procedures to enable appropriate and efficient access to routine health data for research

- Improve transparency and communication around routine health data access and use between data provider organisations and researchers

- Provide better training and support for researchers to use routine health data

These recommendations are described in more detail in a forthcoming PHG Foundation report to be published in July 2017.

---

Contact: Dr Mark Kroese

mark.kroese@phgfoundation.org

PHG Foundation is a health policy think tank with a special focus on how genomics and other emerging health technologies can provide more effective, personalised healthcare

🐦 **@PHGFoundation**
**www.phgfoundation.org**