Linking and sharing routine health data for research in England

Dr Charlotte Warren-Gash
Visiting Fellow PHG Foundation
Background

Healthcare delivery generates increasingly large amounts of routine health data

But are we realising the potential of “big data” to improve health and healthcare systems?
Objectives

- To understand researchers’ experiences of accessing and working with electronic health data on infections in England
- To compare with models of good practice elsewhere in the UK and internationally
- To make recommendations informed by interviews and wider literature to facilitate more effective population health research to be conducted using these datasets
Methods

- We carried out semi-structured interviews with a purposive sample of ten data users between January and March 2017.

- All had experience of using linked EHRs including infection data for research.

- They represented 3 HEIs and ranged in seniority from post-doctoral researchers to professors.
Data used for infection research

- Primary care data
- Secondary care data
- Infection surveillance data
- Cohort data +/- biological samples

Routine data

- Laboratory tests
- Hospital admissions
- GP consultations
- Sentinel surveillance
- Social media data
- Online surveys
Advantages of routine health data

**General**
- Efficiency
- Enhanced power
- Greater phenotypic depth
- Whole populations

**Focussed on infection surveillance-EHR data**
- More complete, specific and detailed diagnoses
- Can answer a wider range of policy-relevant questions
Data users’ experiences of data access

For pre-linked data
• Easier
• ‘Quicker’
• Geographical differences
• Being placed in a data provider organisation e.g. through an honorary contract helps

For new linkages
• Very slow (can take several years)
• Many sequential overlapping access procedures
• Frustrating and challenging
Barriers to data access and use
Cultural barriers

• For some data providers, facilitating secondary research is not a core function

• They may lack capacity or expertise to provide de-identified data for research

• Data access procedures may not be transparent

• There may be unwillingness to share data or concerns over the legitimacy of data sharing
‘With many organisations…their attitude is to pick and choose depending on what research meets their policy objectives and is not in competition with them. There is a lot of ducking and diving and forming relationships’. Data user
Governance barriers

• Information governance requirements for researchers can be disproportionate to risks involved

• Researchers may need support with understanding and negotiating the legal, ethical and governance frameworks and requirements for data access
‘We need to remove barriers and red tape, for example having to get 15 different approvals for the same project, each of which takes three to four weeks’.

Data user
Technical barriers

• Linked datasets are complex: substantial work is needed to clean, transform and manipulate data

• Linkage may be hampered by lack of common identifiers, especially for marginalised groups

• Barriers to interpretation of results include lack of information about the linkage process and data collection context
Examples of data linkage and sharing for research working well

- Ability to link multiple routine datasets
- Streamlined, transparent access procedures
- Good communication between data users and providers
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
Recommendations 2

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

2. Provide better support for researchers working with routine health data.
Acknowledgements

- Dr Mark Kroese, Dr Sobia Raza, Alison Hall (PHG Foundation)
- All data users who agreed to be interviewed