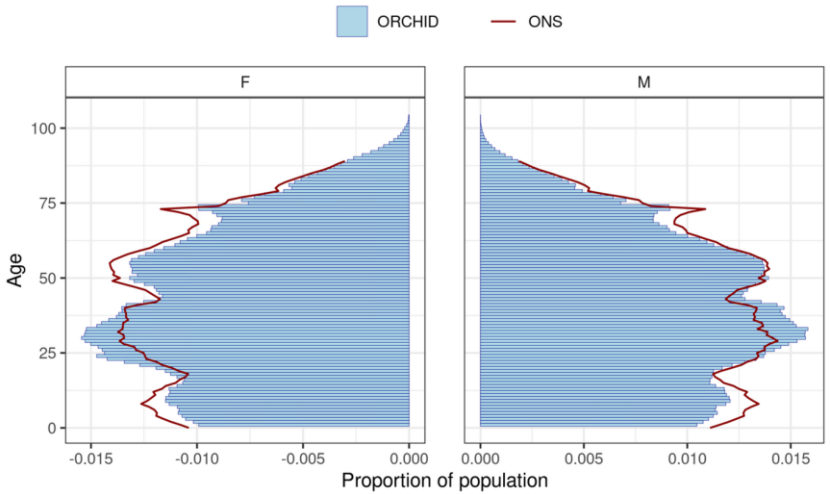
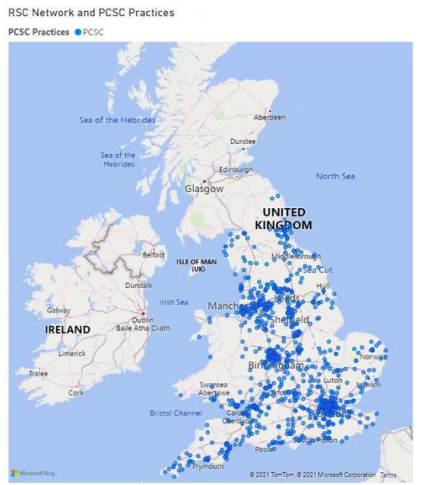
Opportunities and Challenges Associated with Diverse Health Data Linkages: Perspectives of those Working within a Trusted Research Environment

Sneha Anand and Meredith Leston
Clinical Informatics and Health Outcomes Research Group
The Nuffield Department of Primary Care Health Sciences at the University of Oxford

The Oxford-Royal College of General Practices Research and Surveillance Centre (RSC) and its uniqueness

- One of Europe's oldest sentinel systems, working with the UKHSA and its predecessor bodies for 55 years.
- Conducts sentinel surveillance from a nationally representative group of general practices (>1800; ~18 million patients) and provides daily syndromic surveillance data to UKHSA; cohort is split into PCSC and SSGP (pictured).
- >250 RSC practices also collect virology and/or serology specimens to enable real-time disease/ VE surveillance; these samples are also health record-linkable.
- The RSC's data are pseudonymised, linked to hospital and other relevant health datasets/registries (expanded on the next slide) and are held in the Oxford-RCGP Clinical Informatics Digital Hub (ORCHID), a trusted research environment (TRE).
- The RSC's Surveillance report is now published twice weekly, supplemented by real-time online observatories.





Leston, M. & Elson, W. (2022). A Profile of Influenza Vaccine Coverage for 2019-2020: a Database Study of the English Primary Care Sentinel Cohort. Pending publication. Nuffield Department of Primary Care Health Sciences, University of Oxford.

RSC Data Sources and its linkages

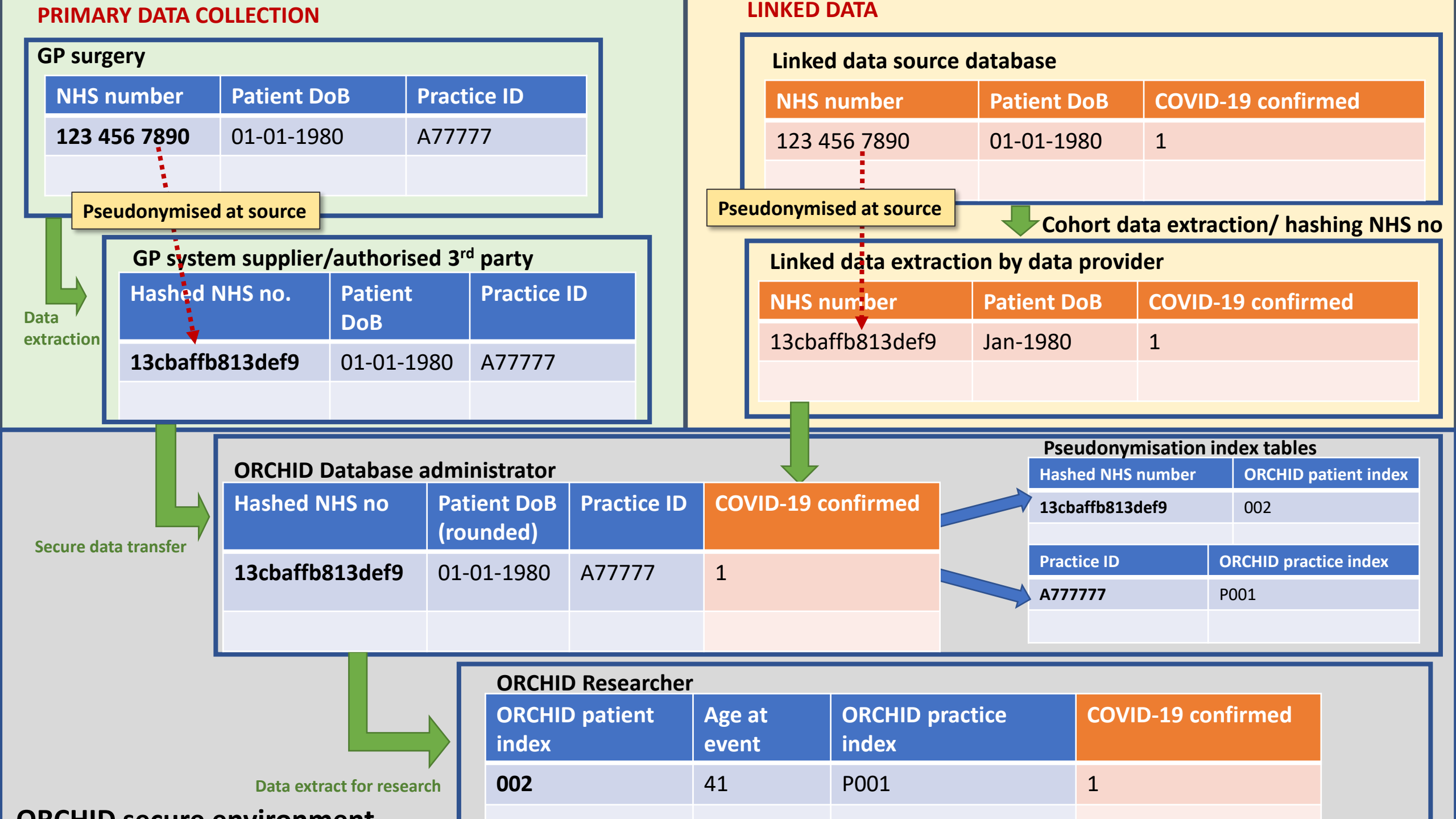
- Oxford-RCGP RSC have the following data sources:

a. Primary care computerised medical records (GP data)
b. Virology, serology and pathology records from UKHSA
c. Secondary care data from NHS Digital
d. Limited consented data directly from patients

Snapshot of datasets received from NHS Digital

NHS (National Health Service) Digital:

- Second Generation Surveillance System (SGSS): laboratory test reports
- National Immunisation Management System (NIMS): COVID-19 vaccination
- Emergency Care Data Set (ECDS)
- Secondary Uses Service (SUS)
- Diagnostic Imaging Dataset
- COVID-19 Hospitalisation in England Surveillance System (CHESS) Dataset-CV19
- NHS 111 Online
- Hospital Episode Statistics (HES; same granularity of supply as for SUS)
- Office of National Statistics (ONS) Mortality
- NHS 111 (free-to-call single nonemergency number medical helpline)
- Cancer Registration Data
- Mental Health Services Data Set


PRIMARY DATA COLLECTION
LINKED DATA
GP surgery
NHS number
Patient DoB
Practice ID
123 456 7890
01-01-1980
A77777
Pseudonymised at source
Linked data source database
NHS number
Patient DoB
COVID-19 confirmed
123 456 7890
01-01-1980
1
Pseudonymised at source
Cohort data extraction/ hashing NHS no
Data extraction
GP system supplier/authorised 3rd party
Hashed NHS no.
Patient DoB
Practice ID
13cbaffb813def9
01-01-1980
A77777
Linked data extraction by data provider
NHS number
Patient DoB
COVID-19 confirmed
13cbaffb813def9
Jan-1980
1
Secure data transfer
ORCHID Database administrator
Hashed NHS no
Patient DoB (rounded)
Practice ID
COVID-19 confirmed
13cbaffb813def9
01-01-1980
A77777
1
Pseudonymisation index tables
Hashed NHS number
ORCHID patient index
13cbaffb813def9
002
Practice ID
ORCHID practice index
A777777
P001
Data extract for research
ORCHID Researcher
ORCHID patient index
Age at event
ORCHID practice index
COVID-19 confirmed
002
41
P001
1
ORCHID secure environment

Royal College of General Practitioners
Research & Surveillance Centre

Oxford–Royal College of General Practitioners
CLINICAL INFORMATICS DIGITAL HUB
ORCHID

UNIVERSITY OF OXFORD

UK Health Security Agency

Scope of data collected

- Sociodemographic data – age, gender, ethnicity, socioeconomic status (SES)
- Lower Super Output Area (LSOA)
- NHS administrative area and region
- Smoking status
- Obesity
- Vaccination history
- Health outcomes data (HES Outpatients, A&E, ICU critical care, death)
- Consultation frequency and attendance
- Comorbidity and Frailty scores

- As specified in DSAs with member practices, onward data linkages are also only permissible when in keeping with SQUIRE (Surveillance, QUality Improvement, Research and Education) purposes
- Limitations on ORCHID data usage or onward linkage by researchers are determined on a user-specific or project-specific basis; in accordance with governance and appropriate permissions
- We respect patient opt-outs : patients that decline to share their data are excluded from any extraction process
- Only non-identifiable data leaves our secure network
- ORCHID and RCGP RSC do not provide licensed datasets or copies of the core datasets and only sub-sets of the data are released to researchers as appropriate for their projects

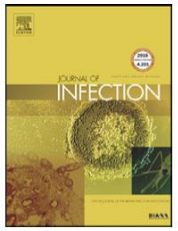
Diverse Data Linkage in Action: Vaccine Effectiveness Surveillance

Journal of Infection 84 (2022) 814–824

Contents lists available at ScienceDirect

Journal of Infection

journal homepage: www.elsevier.com/locate/jinf



Sociodemographic disparities in COVID-19 seroprevalence across England in the Oxford RCGP primary care sentinel network

Heather Whitaker[b,*], Ruby S.M. Tsang[a], Elizabeth Button[a], Nick Andrews[b,c], Rachel Byford[a], Ray Borrow[d], F.D. Richard Hobbs[a], Tim Brooks[f], Gary Howsam[g], Kevin Brown[c], Jack Macartney[a], Charlotte Gower[c], Cecilia Okusi[a], Jacqueline Hewson[e], Julian Sherlock[a], Ezra Linley[d], Manasa Tripathy[a], Ashley D. Otter[e], John Williams[a], Simon Tonge[d], Simon de Lusignan[a], Gayatri Amirthalingam[c]

[a] Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford OX2 6GG, UK
[b] Statistics, Modelling and Economics Department, UK Health Security Agency, 61 Colindale Avenue, London NW9 5EQ, UK
[c] Immunisation and Countermeasures Division, UK Health Security Agency, 61 Colindale Avenue, London NW9 5EQ, UK
[d] Vaccine Evaluation Unit, UK Health Security Agency, Manchester M13 9WL, UK
[e] Diagnostics and Genomics, National Infection Service, Public Health England, Porton Down, Salisbury SP4 0JG, UK
[f] Rare & Imported Pathogens Laboratory, UK Health Security Agency, Porton Down, Salisbury SP4 0JG, UK
[g] Royal College General Practitioners, London NW1 2FB, UK

Royal College of General Practitioners Research & Surveillance Centre

Oxford–Royal College of General Practitioners Clinical Informatics Digital Hub (ORCHID)

University of Oxford

UK Health Security Agency

Diverse Data Linkages: Current Opportunities from a Researcher Perspective

- The UK is a uniquely linkable health ecosystem via NHS numbers; if we can't pursue diverse health data linkages, then it's hard to see how anyone else could
- Massive improvements made recently in coding consistency/ centralised and standardised working amongst clinical systems and healthcare workers – should be leveraged for this agenda
- Potential driver of personalised / precision medicine
- Facilitates discovery via decentralised clinical trials (drug/ treatment/ optimal triage etc.) – cuts costs by not having to follow-up with participants longitudinally on a one-to-one basis
- Provides ample opportunities for patient involvement in research
- Provides contextual information needed to analyse disease patterns at multiple levels of impact/ granularity e.g. national, local, household, demographic etc.
- Provides entire patient arc (e.g. health outcomes onwards from breakthrough infection)
- Identifies unexpected relationships e.g. pollution levels' association with mental health prescriptions
- Facilitates government ambitions for linking health and social care sectors more generally

Royal College of General Practitioners
Research & Surveillance Centre

Oxford–Royal College of General Practitioners
ORCHID Clinical Informatics Digital Hub

University of Oxford

UK Health Security Agency

Diverse Data Linkages: Current Challenges from a Researcher Perspective

- Bad data, bad linkage (issues of codification vs free-text, missing data etc.)
- Interoperability remains poor (between data types, coding languages, nations etc.)
- Linking is an imperfect science prone to cherry-picking/ improper imputation
- Confusion between pseudonymised, anonymised and encryption
- Deep expertise of researchers sometimes limits wider or ‘outside of comfort zone’ linkages
- Data sets often come without clear population denominators – understanding prevalence becomes difficult
- Reidentification remains a problem – not many parameters are needed for a ‘motivated intruder’ to identify a target (e.g. triangulation)
- Disputes over what constitutes a health record (e.g. imprisonment information)
- Inconsistent data access request procedures
- It’s difficult to diagnose exactly where problems in linkages have occurred e.g. duplicate records – which is the correct one?

Royal College of General Practitioners Research & Surveillance Centre

Oxford–Royal College of General Practitioners Clinical Informatics Digital Hub (ORCHID)

University of Oxford

UK Health Security Agency

Diverse Data Linkages: Researcher Recommendations

- 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
- A central repository of code types should be made available e.g. how does SNOMED map onto REED etc?
- The loosening of linkage restrictions that occurred during the COVID-19 pandemic should remain in place
- Data access requests must be made consistent across data types for researchers (e.g. same timelines/ documents)
- Those making data access requests must be absolutely clear on what they need from data custodians/ those responsible for curating data
- Work must be done to identify the datasets that, when triangulated, inadvertently reidentify patients
- There needs to be health data consensus between England, Scotland, NI and Wales
- Terms of consent for data use and share must be defined – how broad should this be?
- GDPR must be more accommodating for international disease data share

Royal College of General Practitioners
Research & Surveillance Centre

Oxford–Royal College of General Practitioners
ORCHID
Clinical Informatics Digital Hub

University of Oxford

UK Health Security Agency

Thank you, any questions?

Sneha.Anand@phc.ox.ac.uk
Meredith.Leston@phc.ox.ac.uk