

Building capacity in data science for MLTC research



The Health Data Research UK
(HDRUK) approach

Professor Elizabeth Sapey

University of Birmingham



UK MLTC Symposium
18 April 2024



Medical
Research
Council



Economic
and Social
Research Council

NIHR | National Institute for
Health and Care Research



Health data and its use to study multiple long-term conditions

Professor Liz Sapey
University of Birmingham



UNIVERSITY OF
BIRMINGHAM



University
of Dundee



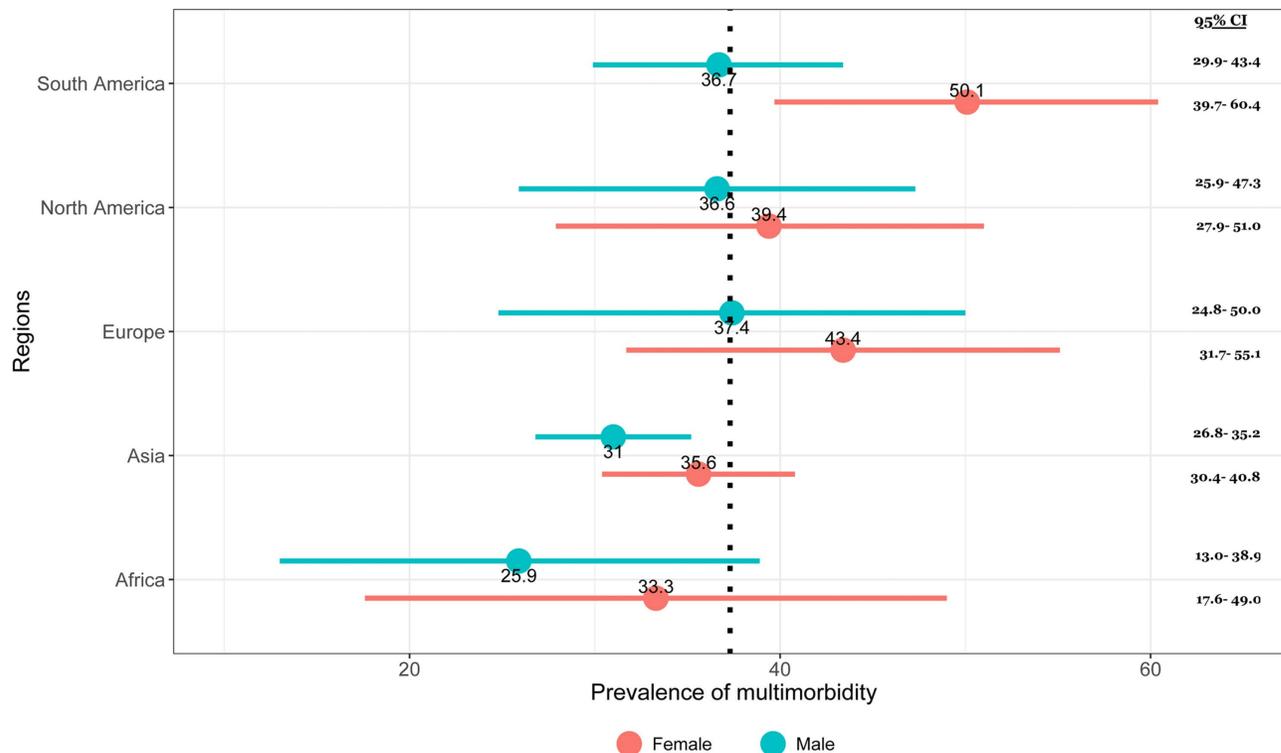
Manchester
Metropolitan
University



The importance of MLTCs



- Increasing prevalence in all communities
- Associated with poor outcomes and increased healthcare use/costs
- Global challenge



eClinicalMedicine
Part of THE LANCET *Discovery Science*

ARTICLES | VOLUME 57, 101860, MARCH 2023

[Download Full Issue](#)

Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis

Saifur Rahman Chowdhury [✉](#) • Dipak Chandra Das [✉](#) • Tachlima Chowdhury Sunna [✉](#) • Joseph Beyene [✉](#) • Ahmed Hossain [✉](#)

Open Access • Published: February 15, 2023 • DOI: <https://doi.org/10.1016/j.eclinm.2023.101860>

The clinical challenge of MLTC care



Multimorbidity: clinical assessment and management

NICE guideline [NG56] Published: 21 September 2016

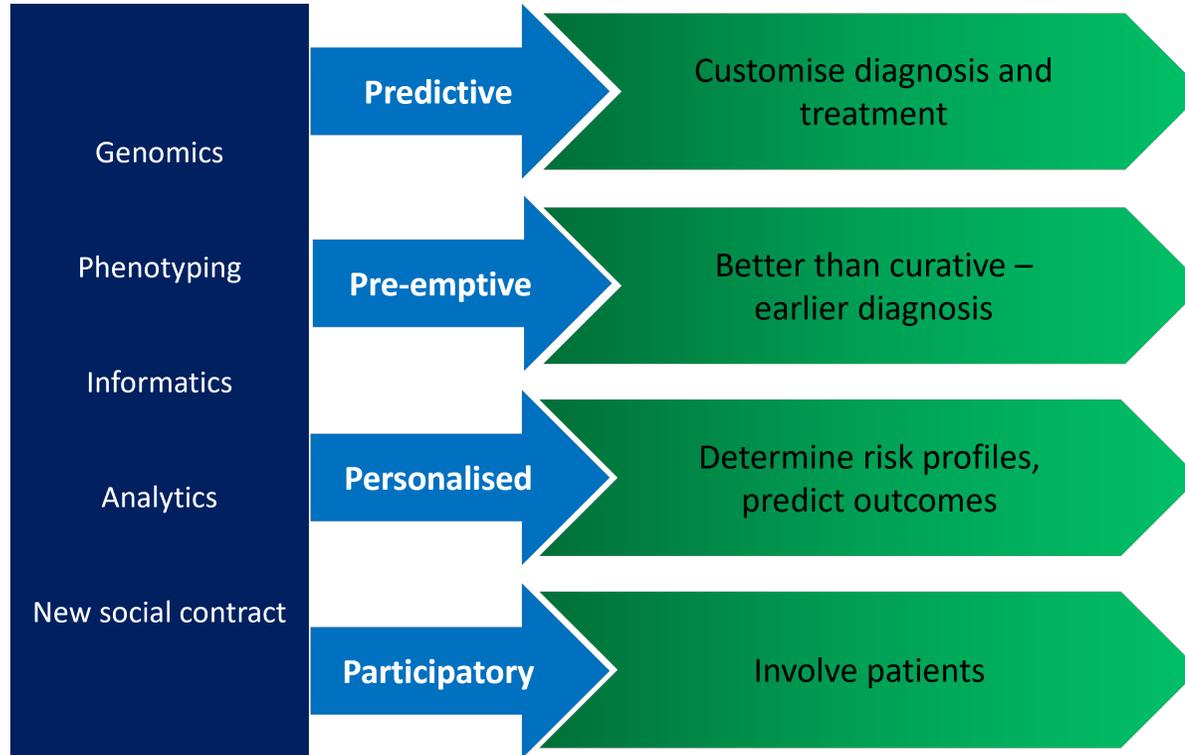


- Multiple OPD attendances in different departments
- Siloed decision making in healthcare “ologies” with potential for harm
- No secondary care ownership of holistic patient care
- Individual patient cost
- Indirect costs – repeated journeys

What could “good” look like?



“4 P” Medicine approach



Joined up approach to:

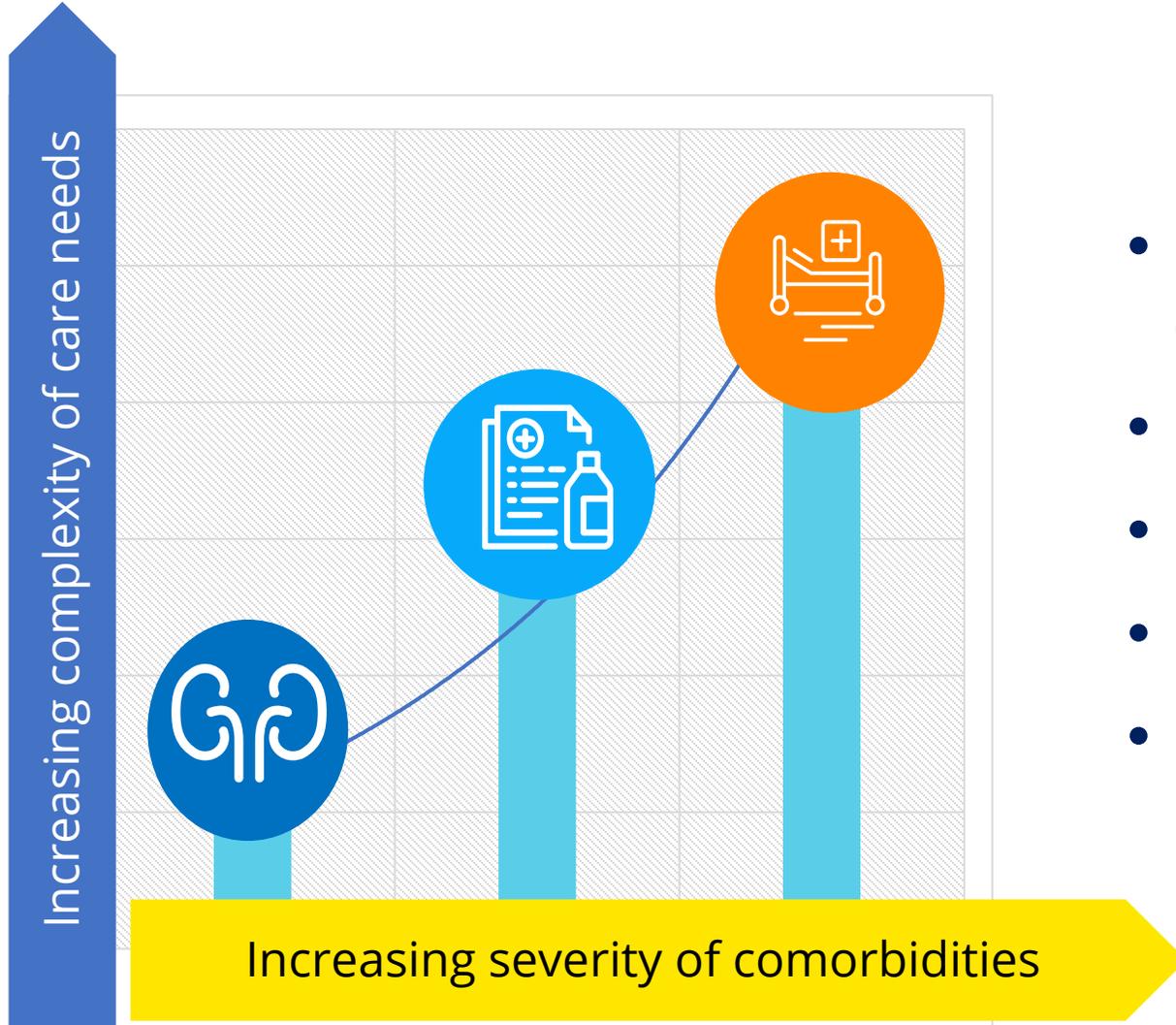
- Predict risk on an individual basis
- Pre-empt each condition before onset
- Personalise treatments based on combination of MLTC present and likely response by that individual.
- Make use of synergistic therapy combinations
- Co-build care pathways which reflect MLTCs and reduce number of health contacts

Current challenges in MLTC research



- Recording of some LTCs is poor in medical coding.
- Not all long-term conditions are equal in terms of impact – but most studies offer a “count”.
- The severity of the LTC is important, yet poorly recorded.
- LTC tend to occur at different points in time - but many studies are cross-sectional.
- Available data does not often reflect the full community - “data poverty” and risk of bias.

The complex nature of MLTC and impact



- Impact of MLTCs likely to depend on:
 - Which combination of MLTC
 - Severity of one or all
 - Symptom burden
 - The complexity of health care needs

- Ethically and CAG approved, secure research database and analytical environment.
- Link patient data at an individual level across community and hospital health care providers.
- Disease and organ agnostic – complete health records.
- Data is near real time, frequently refreshed.
- Build bespoke datasets including synthetic.
- Build bespoke Trusted Research Environments.
- Provide data access under license to NHS, academic, commercial, policy, 3rd sector.
- Provide transparent public oversight through our Data Trust Committee.

Unique
data health
platforms

20 years

Longitudinal data

1.2 million

records

153

UK care hospitals linked
across these records and
all searchable through HDR
UK's Innovation Gateway

Refreshed in near-real-time for
'evergreen' insights

Patient and public involvement



Atkin et al. *Research Involvement and Engagement* (2021) 7:40
<https://doi.org/10.1186/s40900-021-00281-2>

Research Involvement
and Engagement

RESEARCH ARTICLE

Open Access

Perceptions of anonymised data use and awareness of the NHS data opt-out amongst patients, carers and healthcare staff



C. Atkin¹, B. Crosby², K. Dunn³, G. Price⁴, E. Marston⁵, C. C. S. Gallier¹¹, S. Modhwadia¹², J. Attwood¹³, S. Perks¹³, A. K. D. A. Ignatowicz²⁰, H. Fanning²¹, E. Sapey^{22,23,24} and On bel

Abstract

Background: England operates a National Data Opt-Out (NDOO) for research and planning. We hypothesised that public awareness and perceptions of secondary data use, grouping (NHS), academia or commercial. We assessed awareness of staff and the public. We co-developed recommendations to research.

Methods: A patient and public engagement program, co-developed questionnaires and discussion groups regarding anonymised

Results: There were 350 participants in total. Central cancer use, the potential for discrimination and data sharing without their data to be used for NHS research, 85% for academic research, 50% for non-healthcare companies and opinions varied with Questionnaires showed that knowledge of the NDOO was low and 29% of all patients aware of the NDOO.

Recommendations to guide unconsented secondary health patients; data sharing decisions should involve patients/public services with the principles of data minimisation applied. Future health data use, including publicly available lists of projects,

* Correspondence: e.sapey@bham.ac.uk

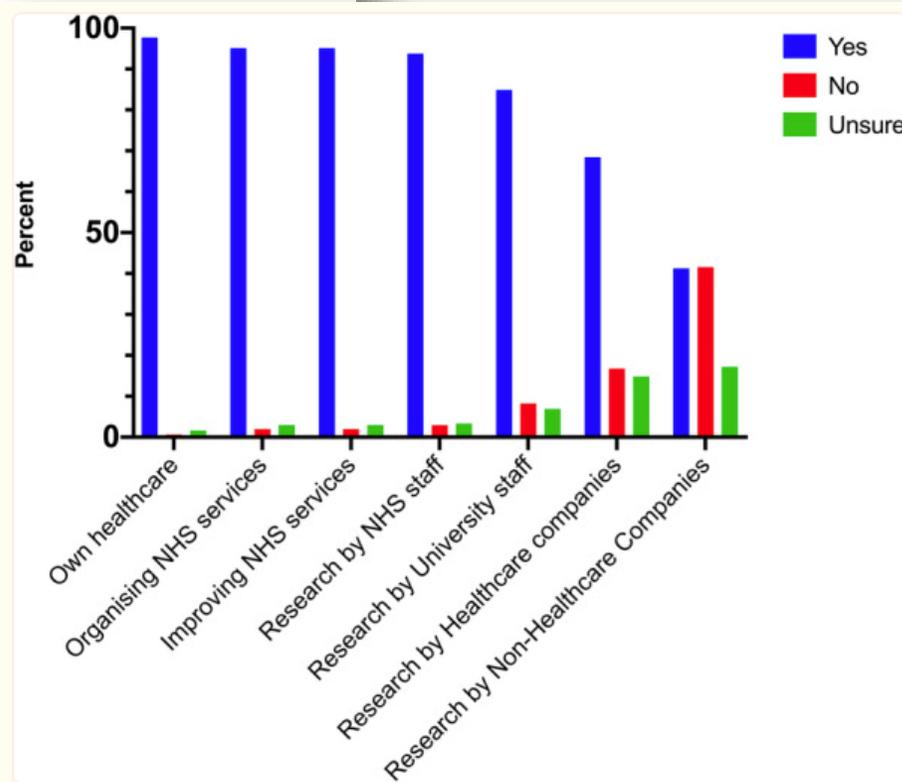
²²POKIER, HDR-UK Health Data Research Hub in Acute Care, Birmingham Acute Care Research Group, Institute of Inflammation and Ageing, University of Birmingham, Birmingham B15 2GW, UK

²³Department of Acute Medicine, University Hospitals Birmingham NHS Foundation Trust, Birmingham B15 2GW, UK

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



- Co-developed protocol.
- Local and national data opt out.
- “Data Trust Committee” review all data access requests.
- Weigh up public benefit versus risk.
- DTC decision is binding.
- Supported >90 data requests since Sept 2020.

Data availability



Attendance Reason Operations/ Procedure



Medications / Allergies



Vital Signs



Laboratory Tests



Imaging Reports



Complications / Diagnosis



Severity



Patient Transfers



Final Outcome



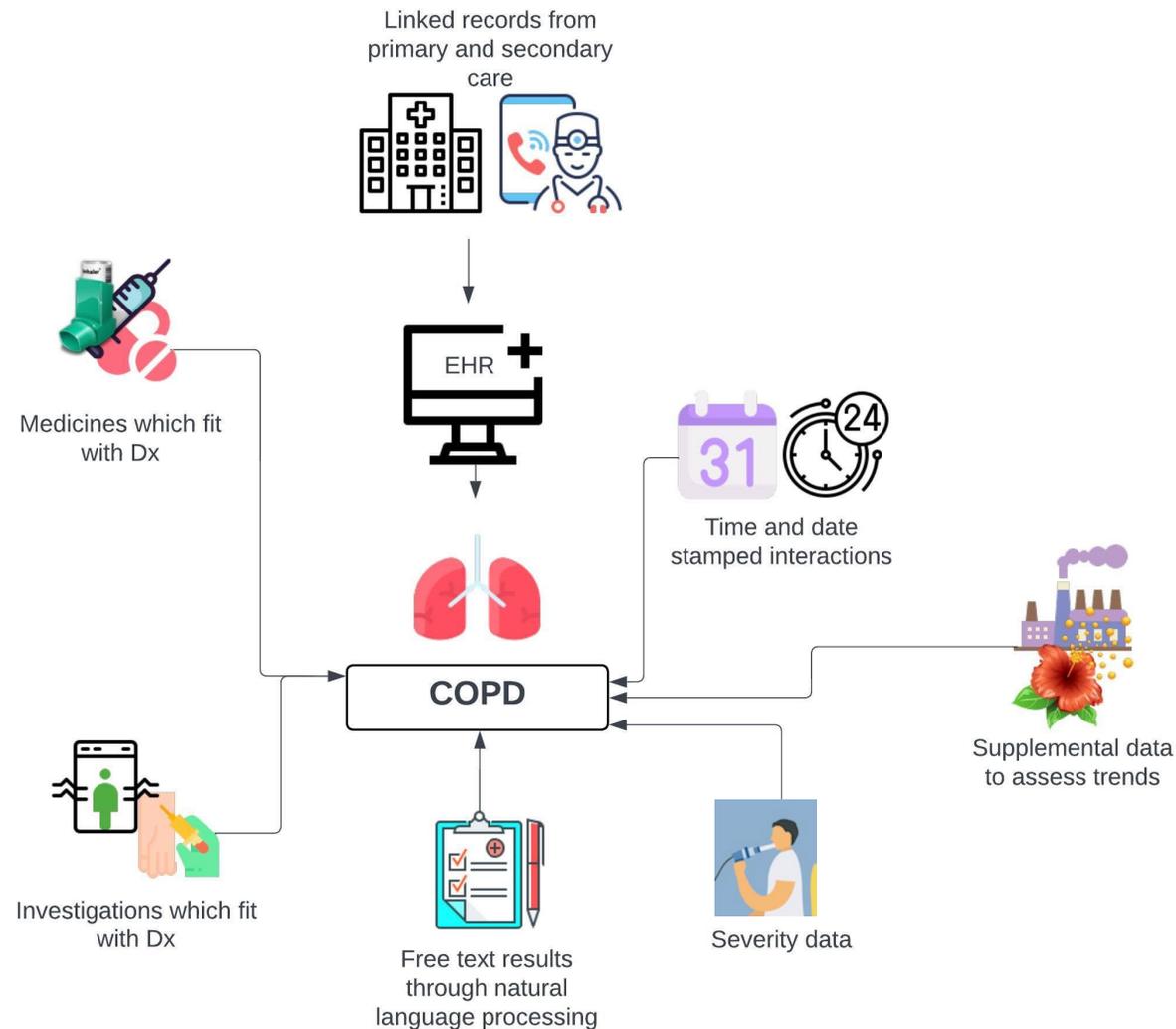
Outpatient Care



**Synthetic data -
including images**

- Map data across coding systems (ICD to SNOMED CT).
- Real world health data and **synthetic**.
- >20 years longitudinal data.
- Open applications for data access.
- **All data access under license.**
- Data staged in TRE, reducing data egress.

Building a clearer picture of a diagnosis ADMISSION



HDRUK Phenotype Library

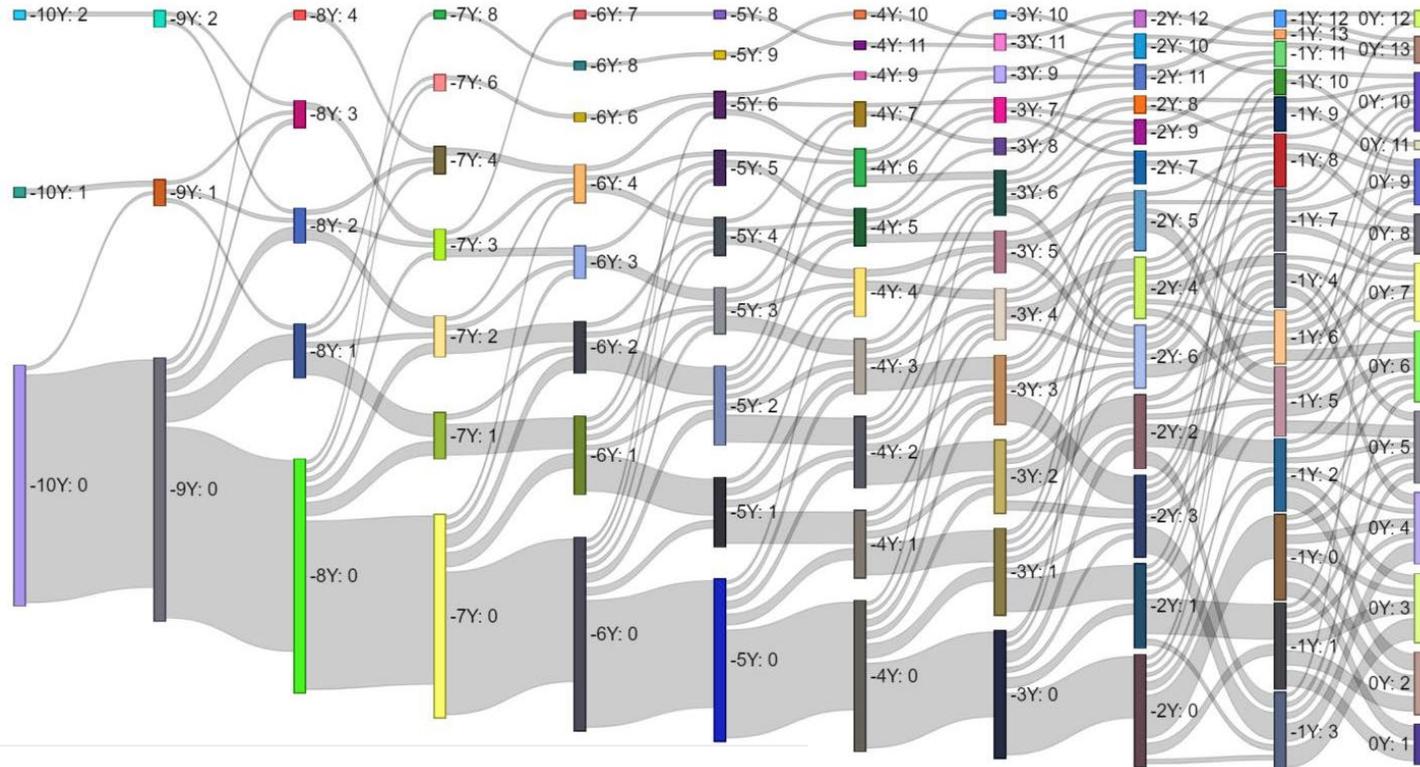
The HDR UK Phenotype Library is a comprehensive, open access resource providing the research community with information, tools and phenotyping algorithms for UK electronic health records.

Search our Phenotype Library

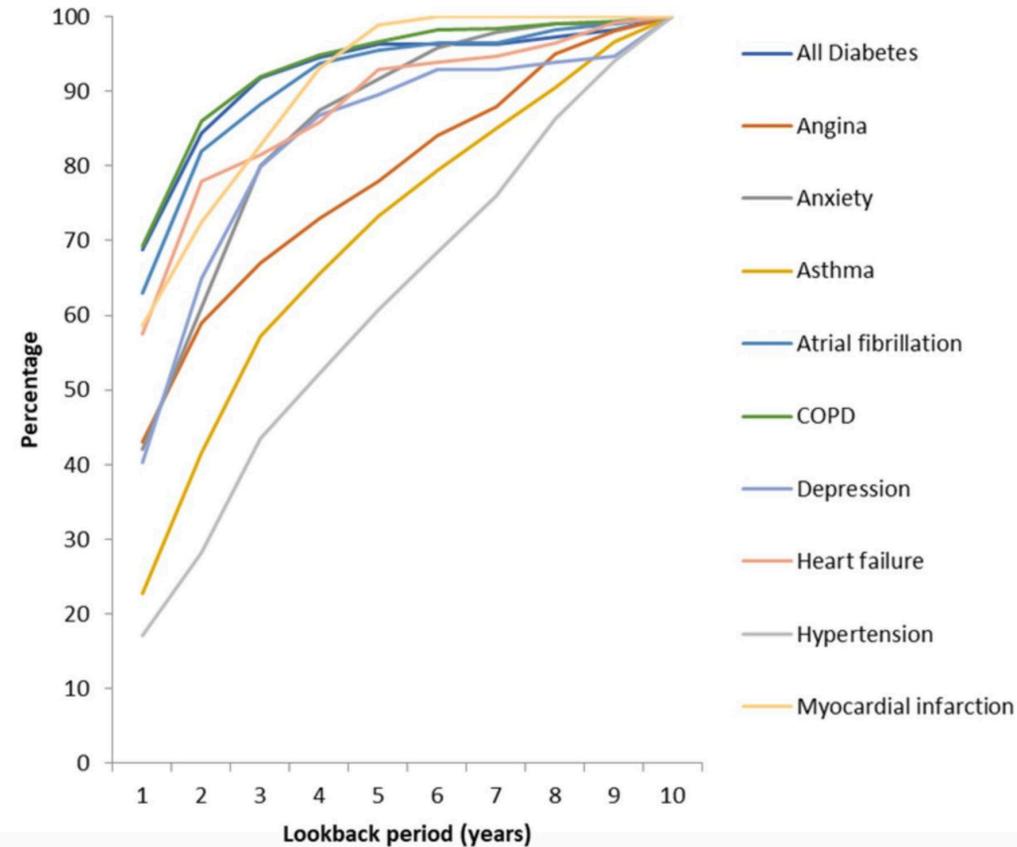
1,090
phenotypes

2,123
codelists

Use case



Percentage of all patients with top 10 grouped conditions per lookback period



Open access Original research

BMJ Open How far back do we need to look to capture diagnoses in electronic health records? A retrospective observational study of hospital electronic health record data

Jadene Lewis,^{1,2} Felicity Evison,^{1,2} Rominique Doal,^{1,2} Joanne Field,³ Suzy Gallier,^{1,2} Steve Harris,^{4,5} Peta le Roux,³ Mohammed Osman,^{6,7} Chris Plummer,^{3,7} Elizabeth Sapey,^{1,8} Mervyn Singer,^{4,9} Avan A Sayer,^{6,7} Miles D Witham,^{6,7} The ADMISSION Research Collaborative

Building opportunities through NHSE SDEs



East Midlands
SECURE DATA ENVIRONMENT

West Midlands
SECURE DATA ENVIRONMENT

Kent, Medway & Sussex
SECURE DATA ENVIRONMENT

North West
SECURE DATA ENVIRONMENT

Thames Valley & Surrey
SECURE DATA ENVIRONMENT

Wessex
SECURE DATA ENVIRONMENT

London
SECURE DATA ENVIRONMENT

North East and North Cumbria
SECURE DATA ENVIRONMENT

East of England
SECURE DATA ENVIRONMENT

Connected Yorkshire
SECURE DATA ENVIRONMENT

Great Western
SECURE DATA ENVIRONMENT

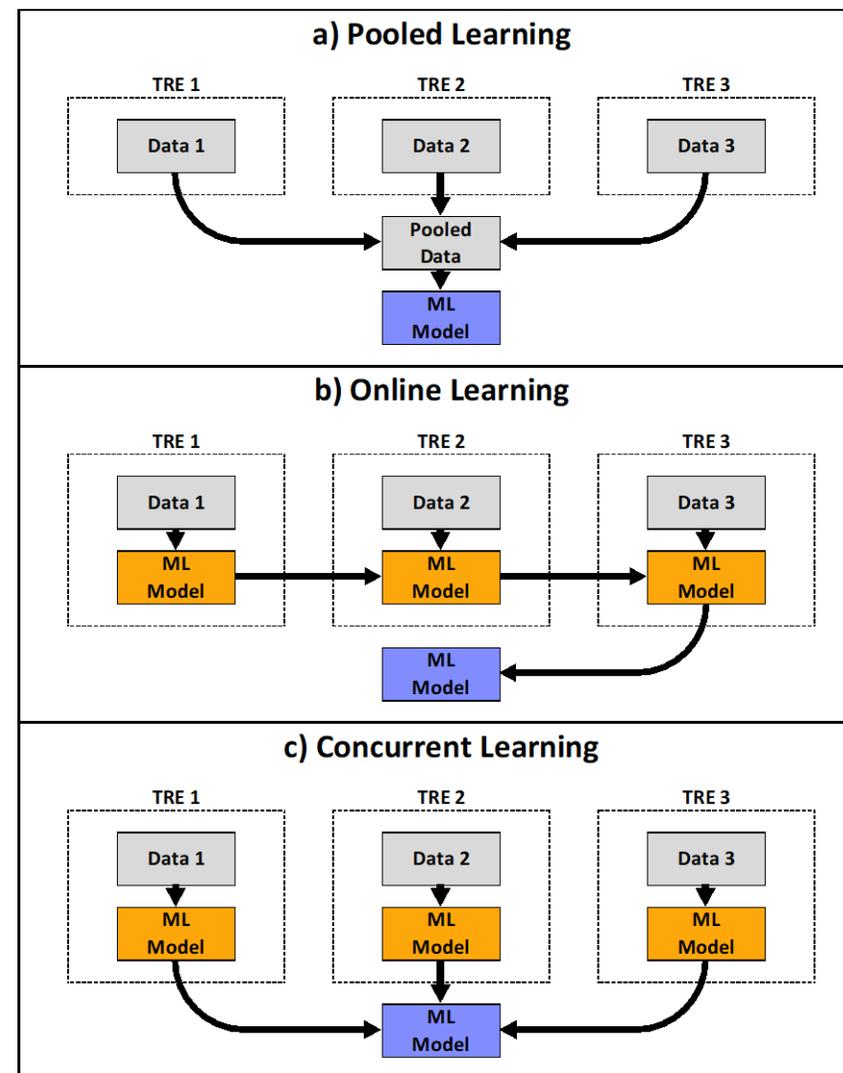


- Funded in 2022
- Processes based on PIONEER
- Ethics and CAG approval to link data from >830 health and care providers
- Even better representation of patient population

Federated analytics

- Data remains with Data Controller.
- Analysts do not “see” raw data.
- Analytical codes moves to TRE.
- Outputs are moved out of TRE.
- Requires
 - Data Controller to manage all data cleansing.
 - Interoperability.
 - Clear metadata to enable building of code.

Testing federated analytics across secure data environments using differing statistical approaches on cross-disciplinary data

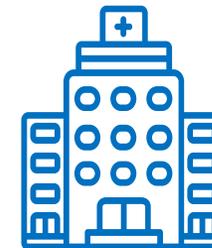
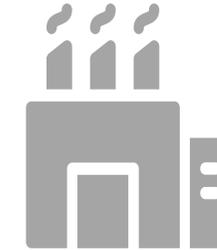
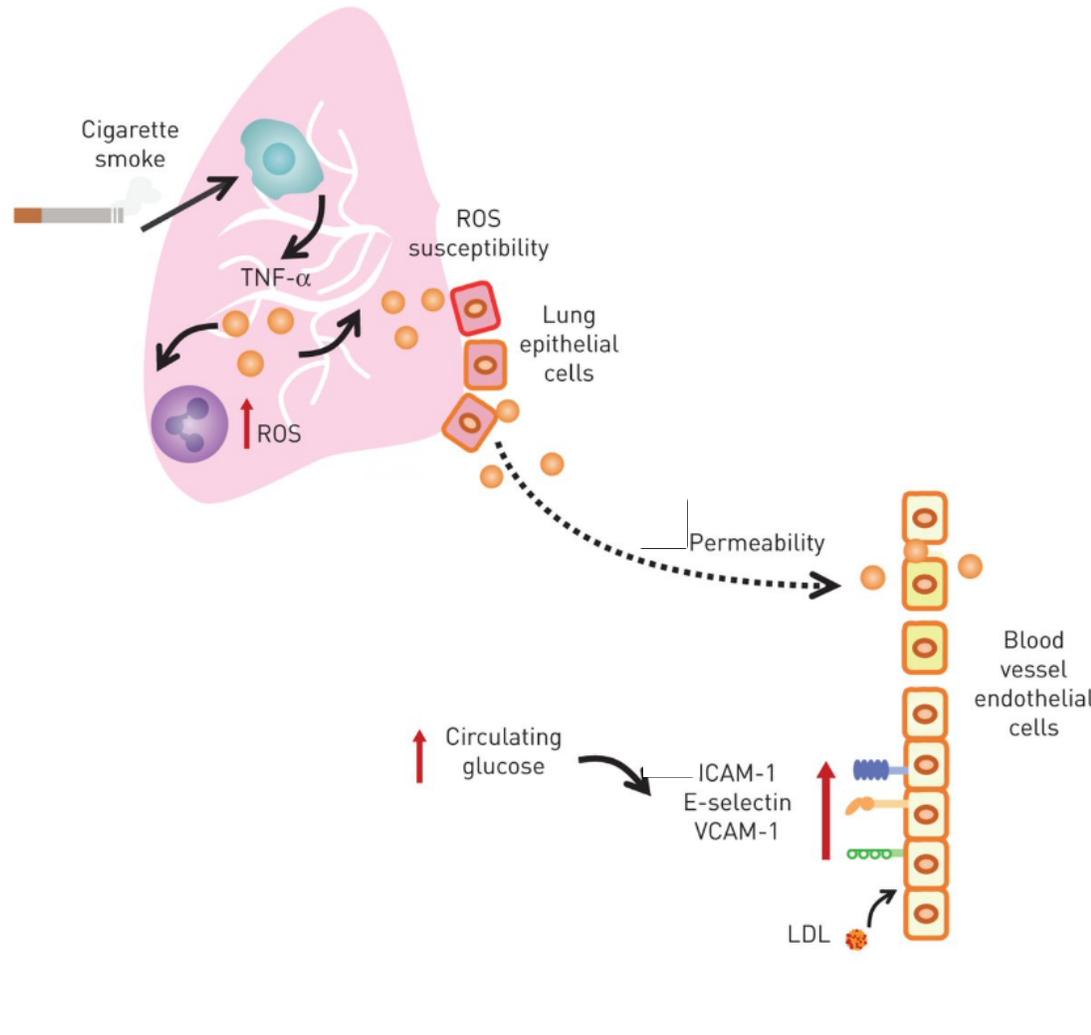




Next steps in health data science: Big data meets little data

- Failure rate for early phase translational studies is high.
- Often translational studies include small numbers – how can data science help?
- Adding translational data to health data to increase opportunities for learning to improve stratified and then personalized medicine

What's next? Potential mechanisms for MLTC





Thank you

Questions?

