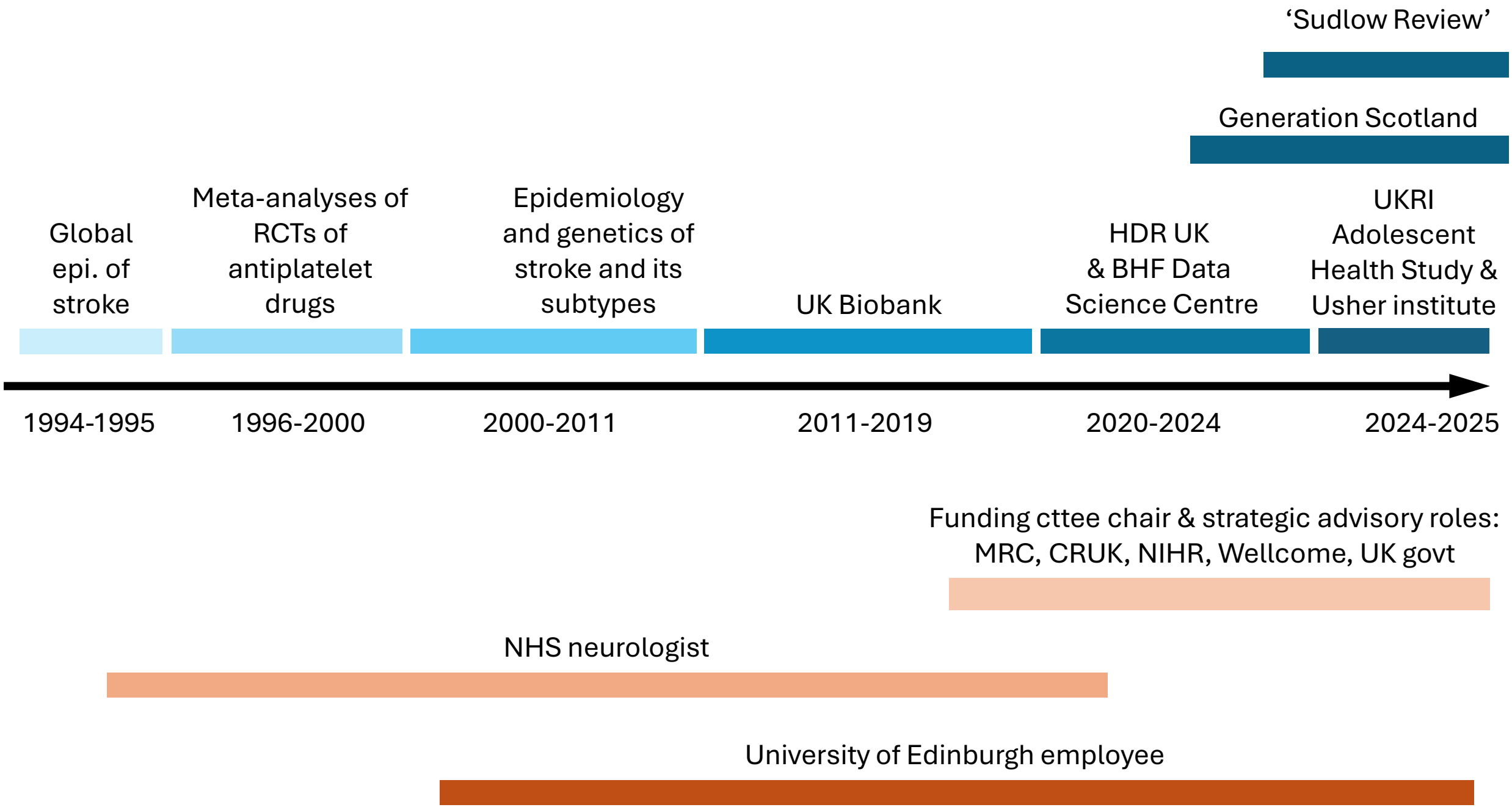
The power of ‘big data’:

perspectives from a neurologist turned population health scientist / data geek / open science resource junky / policy advisor

Cathie Sudlow

Director of Usher Institute and School of Population Health Sciences, University of Edinburgh

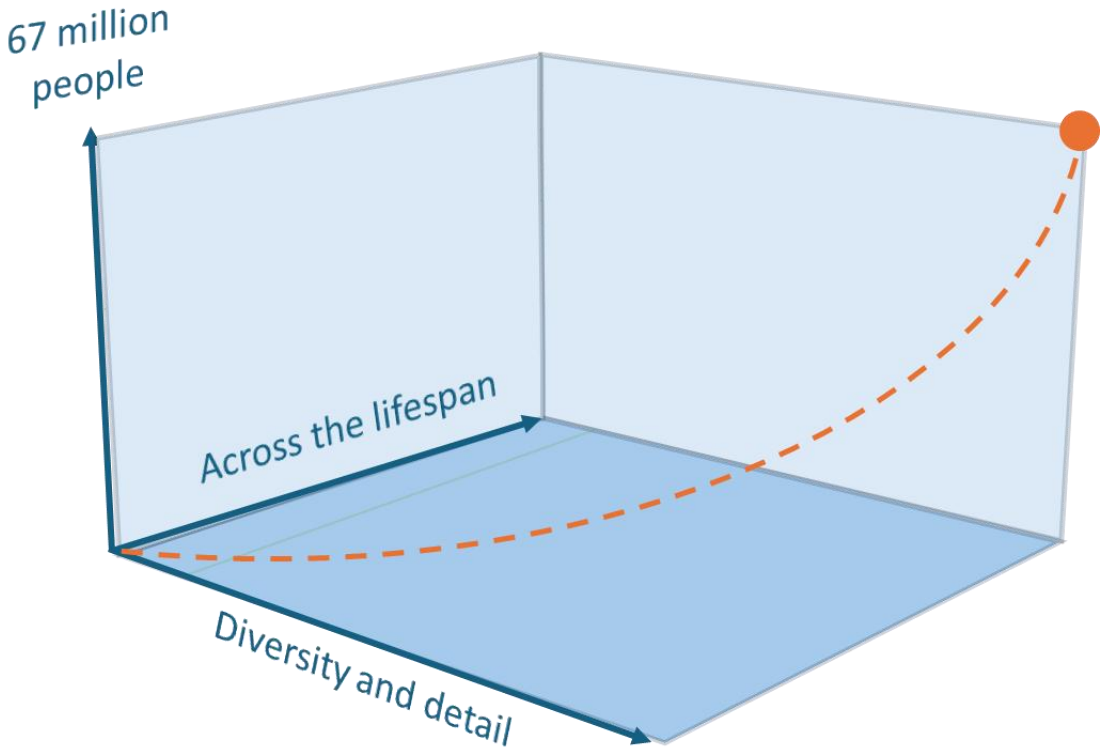
Director of UKRI Adolescent Health Study


'Sudlow Review'
Generation Scotland
Global epi. of stroke
Meta-analyses of RCTs of antiplatelet drugs
Epidemiology and genetics of stroke and its subtypes
UK Biobank
HDR UK & BHF Data Science Centre
UKRI Adolescent Health Study & Usher institute
1994-1995
1996-2000
2000-2011
2011-2019
2020-2024
2024-2025
Funding cttee chair & strategic advisory roles: MRC, CRUK, NIHR, Wellcome, UK govt
NHS neurologist
University of Edinburgh employee

Enabling UK-wide studies to accelerate research that improves lives

British Heart Foundation Data Science Centre
Led by Health Data Research UK

HDRUK
Health Data Research UK



https://www.hdruk.ac.uk/helping-with-health-data/the-sudlow-review/

‘Sudlow Review’: key points

- Health data are used safely every day to improve people’s health and lives
- People overwhelmingly support use of their health data to benefit themselves and others
- The UK has abundant sources of health data, from the NHS and beyond
- Transformational insights come from linking different sources of data together
- But there are many obstacles and delays to access, link and analyse health data
- These barriers arise from the UK’s complex and inefficient systems for managing and accessing data
- They prevent or delay crucial analysis and research about health conditions affecting millions of people across the UK. We are letting people and their families down.
- We need to recognise the UK’s health data as **critical national infrastructure**

Office for Life Sciences summary of 'Sudlow Review': UK has structural advantages in data, with world-leading companies and abundant, rich, diverse data sets but we currently don't make enough of these

Many rich sources of health and care data

UK is home to world-leading companies, academic institutions, and research programmes

\+
Our
Future
Health

biobank uk
Improving the health of future generations

HDRUK
Health Data Research UK

RECOVERY

But we cannot make the most of these data assets

Poor interconnectivity between datasets

Complex and inefficient systems for managing and accessing health data

Huge untapped potential of UK health data, which can...

Support care for each of us

Inform an intelligent healthcare system that predicts and responds to demands

Support the planning and delivery of equitable care locally, regionally, and nationally

Support a thriving research and innovation landscape

But only with an ambitious and joined-up data strategy

Class health data as critical national infrastructure

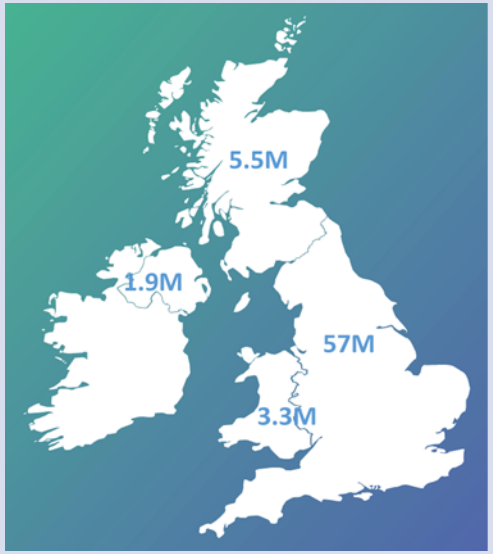
Establish a national health data service

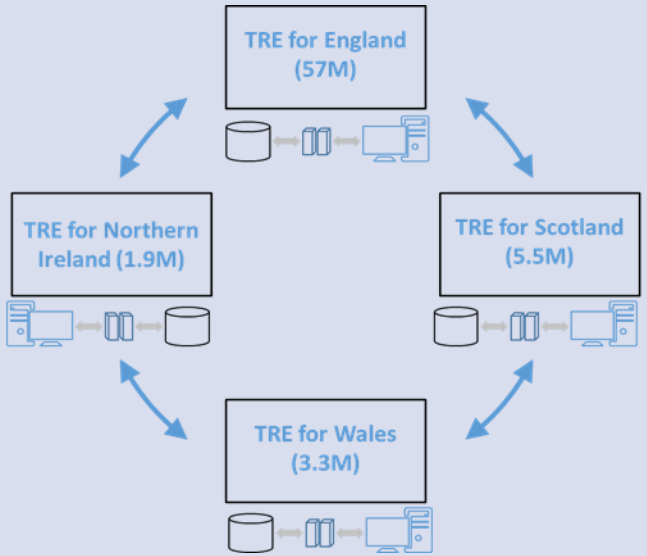
Engage with public and professionals

Establish UK-wide processes and standards

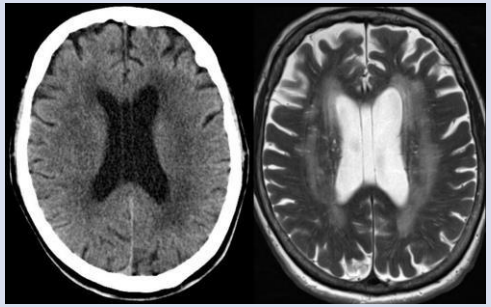
Health-relevant data across the UK

NHS: NATIONALLY COLLATED, POPULATION-WIDE DATA





NHS: LOCAL & REGIONAL DATA



NON-NHS: HEALTH-RELEVANT ADMINISTRATIVE DATA

Census, education, disability, income….

ADR UK

Research Data Scotland

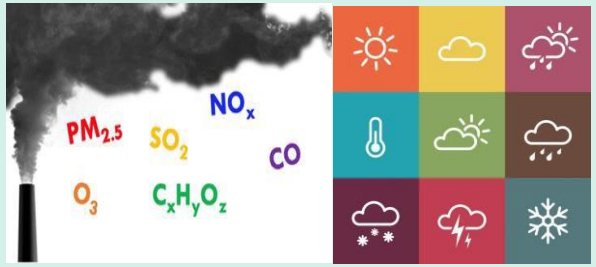
SAIL DATABANK

Northern Ireland Statistics & Research Agency

Office for National Statistics

PERSONAL MONITORING DATA

ENVIRONMENTAL DATA



RESEARCH GENERATED DATA

Cohorts, biobanks, clinical trials, e.g.

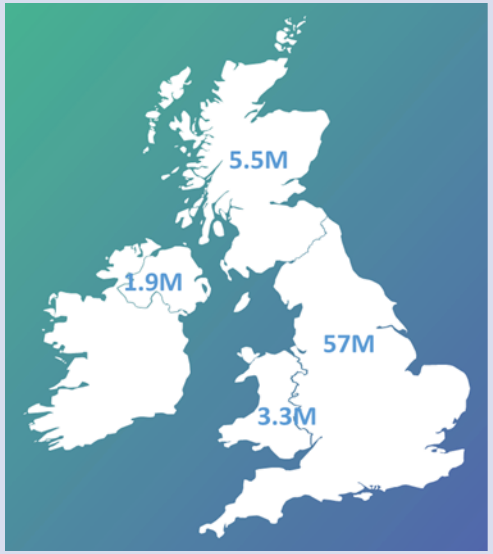
biobank uk

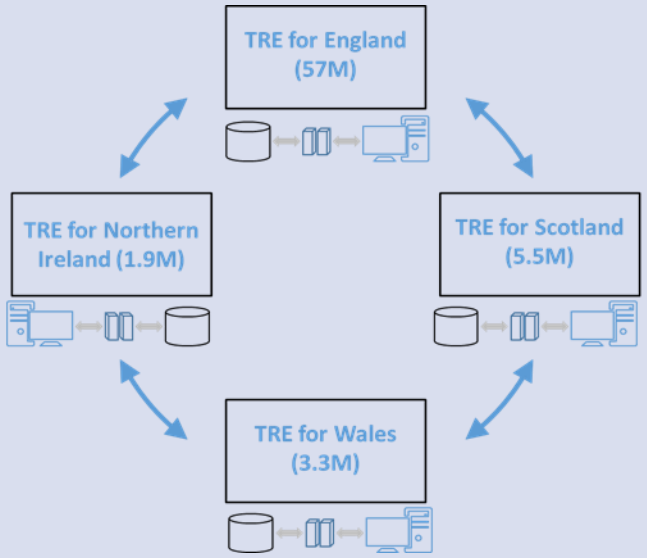
+ Our Future Health

RECOVERY
Randomised Evaluation of COVID-19 Therapy

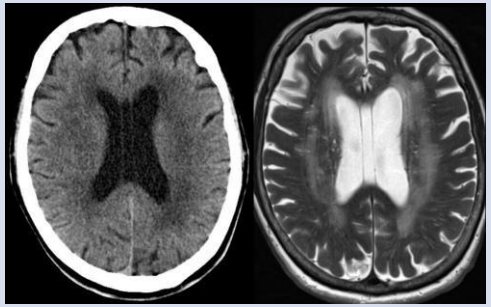
Health-relevant data across the UK

NHS: NATIONALLY COLLATED, POPULATION-WIDE DATA





NHS: LOCAL & REGIONAL DATA



NON-NHS: HEALTH-RELEVANT ADMINISTRATIVE DATA

Census, education, disability, income….

ADR UK

Research Data Scotland

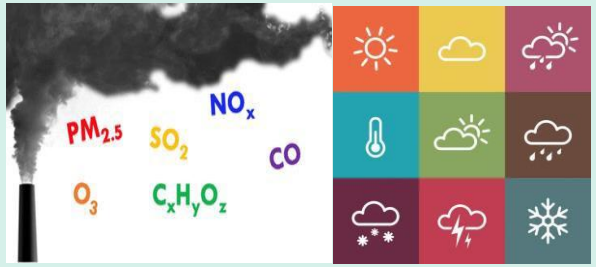
SAIL DATABANK

Northern Ireland Statistics & Research Agency

Office for National Statistics

PERSONAL MONITORING DATA

ENVIRONMENTAL DATA



RESEARCH GENERATED DATA

Cohorts, biobanks, clinical trials, e.g.

biobank uk

+ Our Future Health

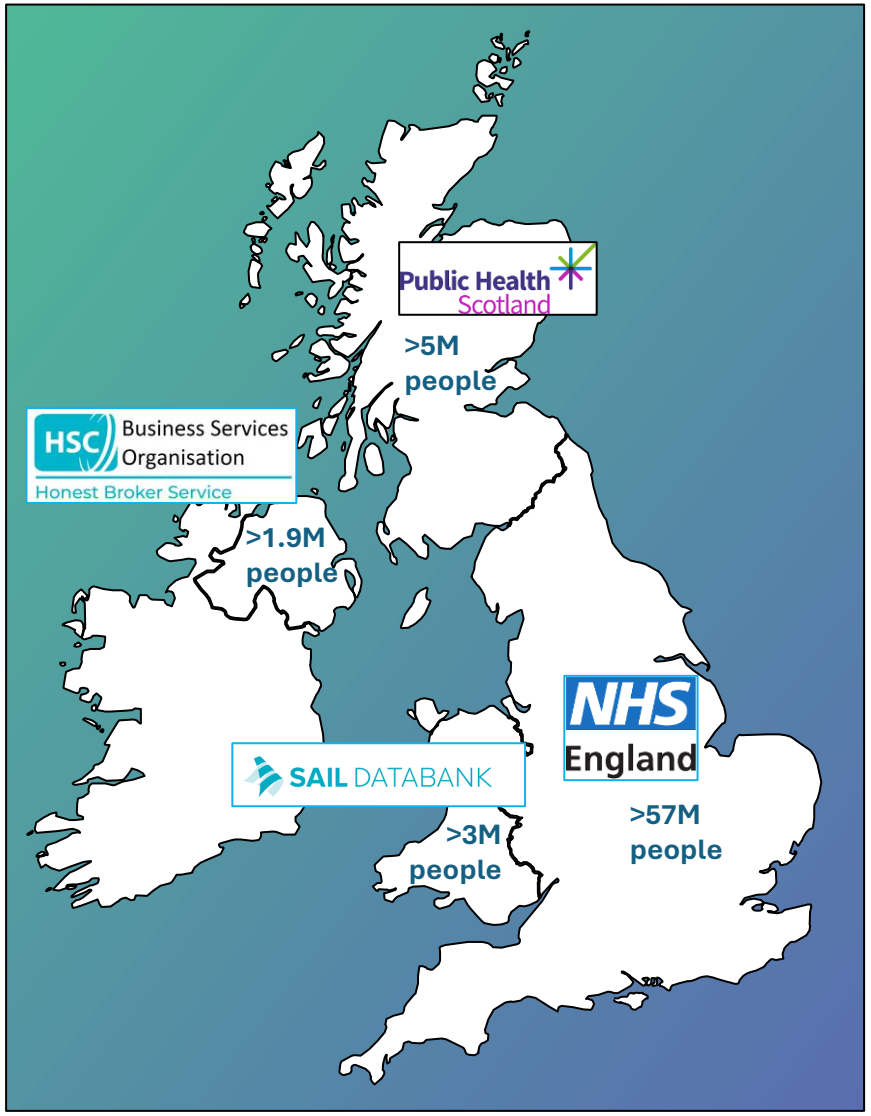
RECOVERY
Randomised Evaluation of COVID-19 Therapy

Uniting the four nations' routinely collected health records for research insights from a population of >67 million people

British Heart Foundation Data Science Centre
Led by Health Data Research UK

HDRUK
Health Data Research UK

- Secure access to multiple linked datasets
  - Primary care
  - Hospital
  - Deaths
  - Medications
  - Specialist registries / audits
  - COVID-19 vaccination and test data
  - Climate/pollution data
- Enables whole population research on **population of >67 million people**:
  - Statistically powerful
  - Comprehensive information on characteristics and health outcomes
  - Including all age groups, ethnicities, geographic locations, socioeconomic, health and personal characteristics
  - Datasets updated regularly



Research Activity

90+ projects

80+ approved researchers

50+ NHS and academic organisations

40+ publications and preprints

Multiple policies

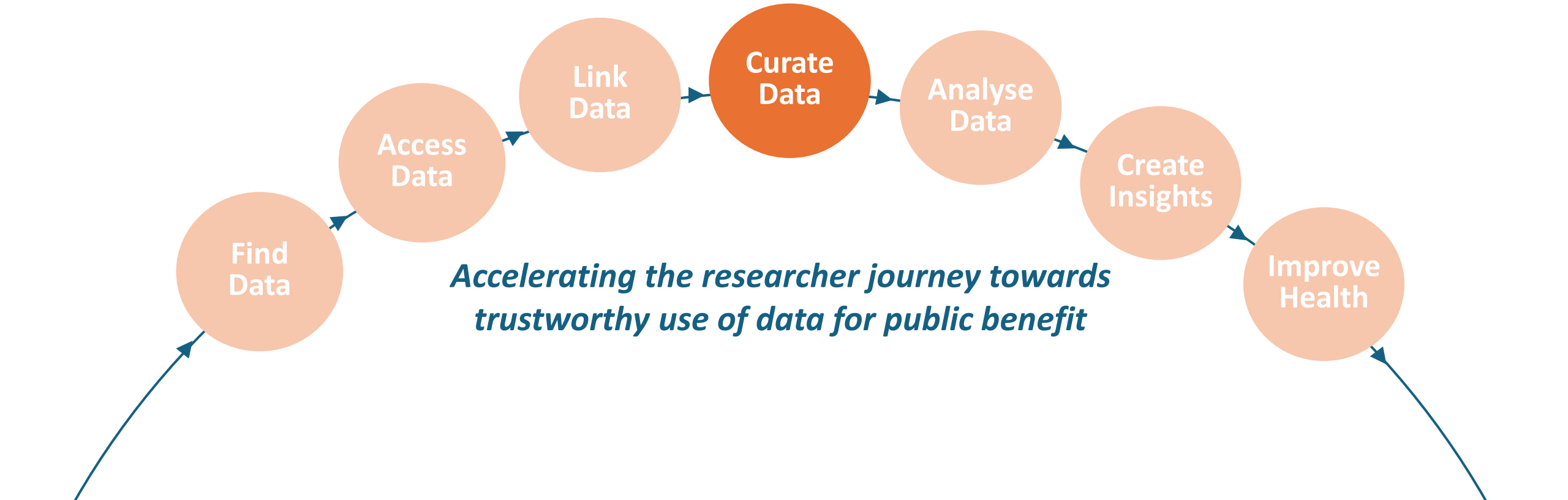
Reproducible data curation pipelines to accelerate researcher productivity

British Heart Foundation Data Science Centre
Led by Health Data Research UK

HDRUK
Health Data Research UK

"**80%** of the work for data science with NHS records is spent on **data preparation**"

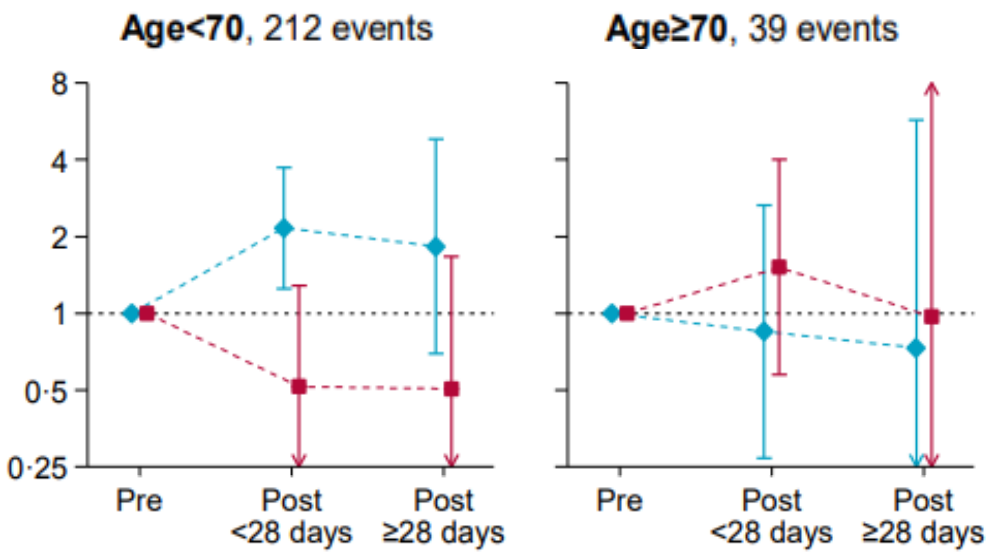
Goldacre Review (2022). Better, Broader, Safer: Using health data for research and analysis.



Rare adverse effects of COVID-19 vaccines among 46 million adults in England

Fully adjusted hazard ratios (& 95% CIs):
pre vaccination (HR=1), 0-28 days, ≥28 days)

Intracranial venous thrombosis (ICVT)

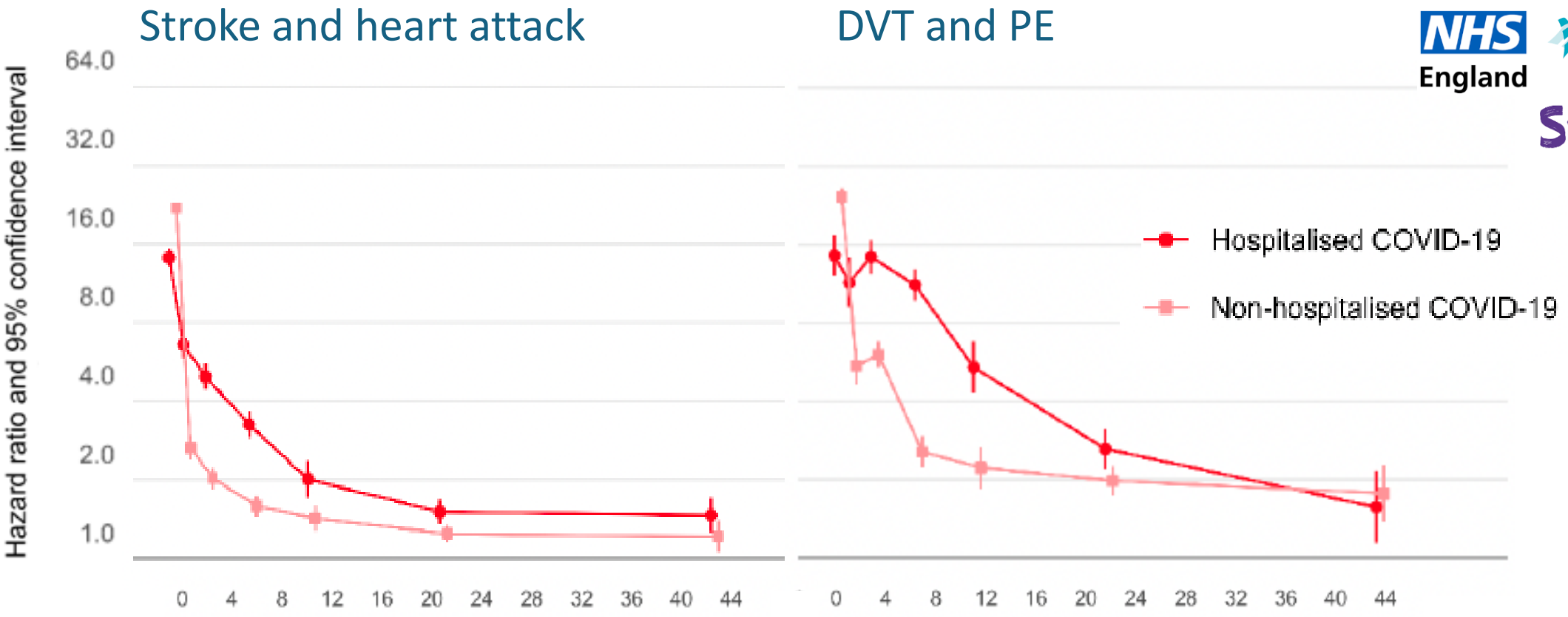


ChAdOx1-S Oxford AZ vaccine
BNT162b2 Pfizer vaccine

- Before the pandemic background rate of up to 3 cases of these rare brain clots per million people per month
- Analyses suggest around 2x risk with Oxford AZ vaccine in those aged <70 years
- So, might expect rate of rare brain clots to double to 6 cases per million in the month post vaccination

Whiteley W et al. PLoS Med 2022

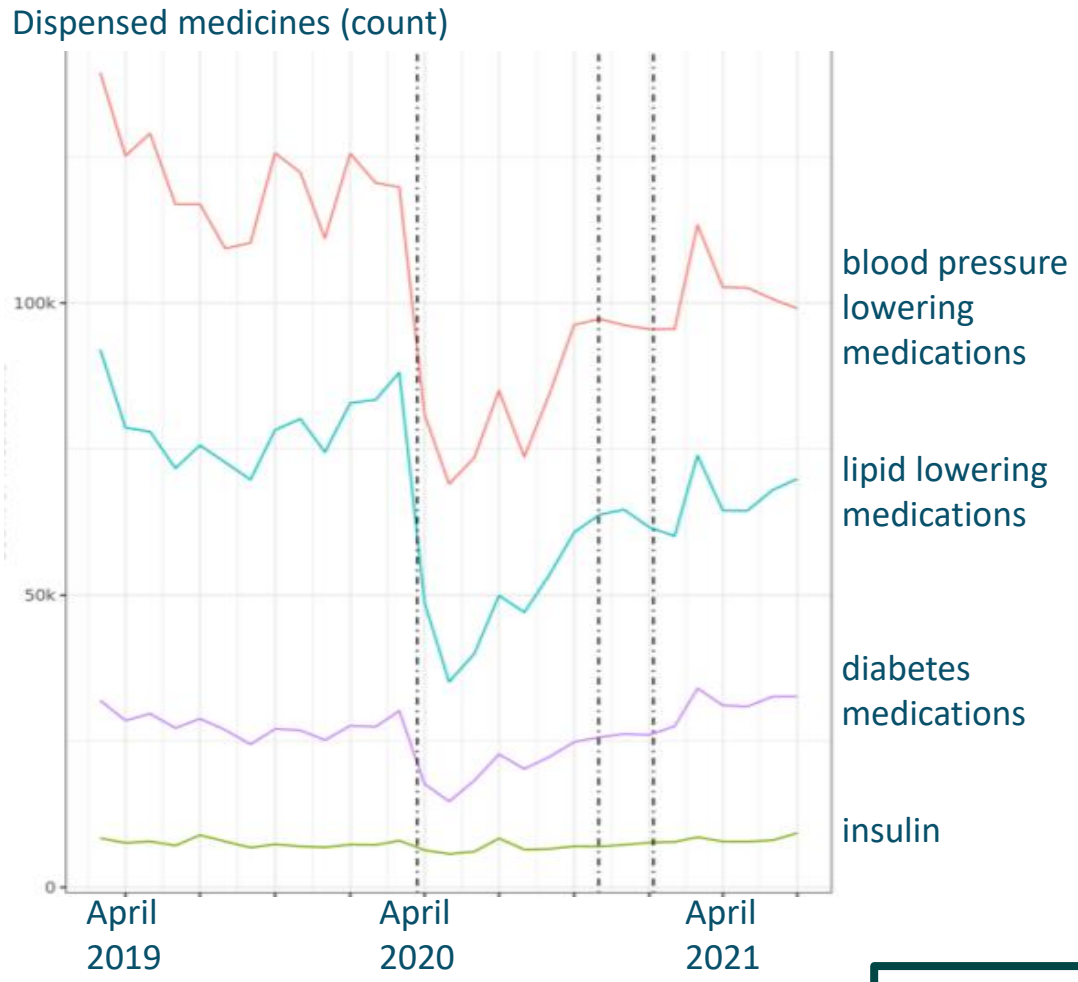
COVID-19 infection & risk of major vascular events among 48 million adults in England and Wales



- Very high early excess relative risk of both venous and arterial events
- Risk falls over time but stays elevated to 40+ weeks post infection
- Relative increased risk falls more slowly for venous compared with arterial events

Knight R et al. Circulation 2022

Linking national medicines data to other national health data sources to understand medicines use across the UK



- New dispenses of antihypertensives, lipid lowering and diabetes medicines (not insulin) fell markedly during 2020-2021 compared with pre-pandemic.
- > 490,000 fewer people than expected started antihypertensive treatment in England, Scotland and Wales from March 2020–May 2021.
- Under-treatment of hypertension predicted to cause
  - > 13,500 additional vascular events
  - > 2,200 additional heart attacks
  - > 3,400 additional strokes

Dale et al. Nature Medicine 2023

bmjmedicine

Check for updates

Use of sodium valproate and other antiseizure drug treatments in England and Wales: quantitative analysis of nationwide linked electronic health records

Dale et al. BMJ Medicine 2024

Linking different sources of national health data in all four UK nations to study under-vaccination against COVID-19 among >64 million people

COALESCE Consortium Lancet 2024

Nation	Under-vaccinated in June 2022
England	46%
Northern Ireland	50%
Scotland	35%
Wales	33%

Risk of getting severe COVID-19 increased in under-vaccinated people by:

- more than 60% in 5-15 year-olds
- almost 50% in 16-74 year-olds
- more than 2 times in >75 year-olds

More likely to be under-vaccinated if

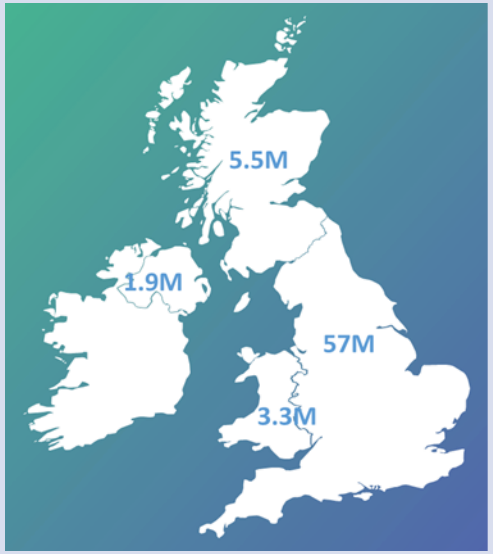
- Younger
- More deprived
- Non-white ethnicity
- Fewer underlying health conditions

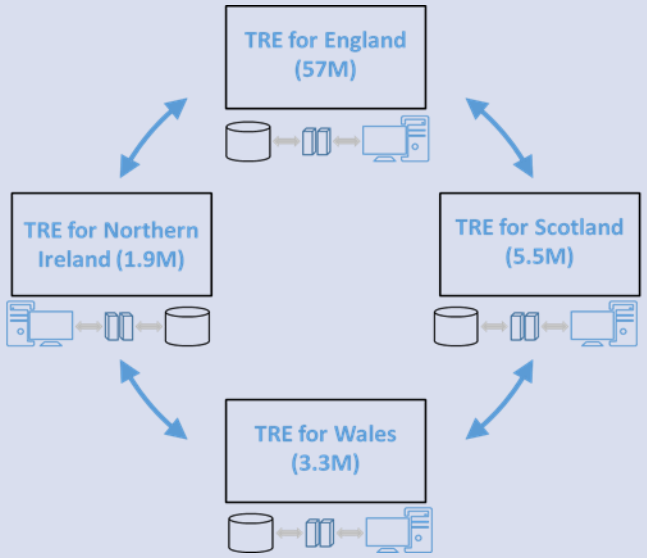
- 7000 severe COVID-19 outcomes over 4 months of follow-up might have been prevented with optimal vaccine coverage
- >70% of these in > 75 year-olds

British Heart Foundation Data Science Centre
Led by Health Data Research UK
BREATHE Health Data Research Hub ASSOCIATED PROJECT
HSC Business Services Organisation Honest Broker Service
SAIL DATABANK
Public Health Scotland
NHS England
THE UNIVERSITY of EDINBURGH
usher institute
UNIVERSITY OF CAMBRIDGE
University of Strathclyde
Swansea University Prifysgol Abertawe
University of BRISTOL
UNIVERSITY OF LEICESTER
UNIVERSITY OF ABERDEEN
UCL

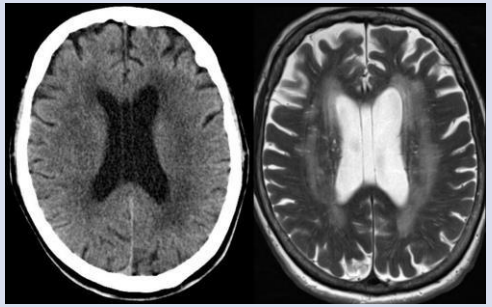
Health-relevant data across the UK

NHS: NATIONALLY COLLATED, POPULATION-WIDE DATA





NHS: LOCAL & REGIONAL DATA



NON-NHS: HEALTH-RELEVANT ADMINISTRATIVE DATA

Census, education, disability, income….

ADR UK

Research Data Scotland

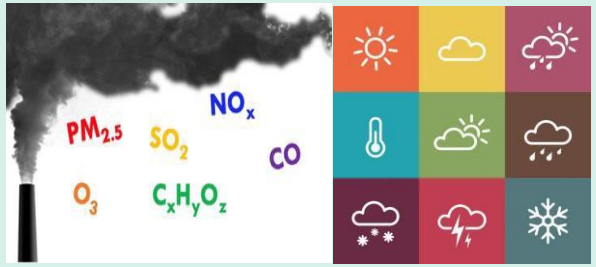
SAIL DATABANK

Northern Ireland Statistics & Research Agency

Office for National Statistics

PERSONAL MONITORING DATA

ENVIRONMENTAL DATA



RESEARCH GENERATED DATA

Cohorts, biobanks, clinical trials, e.g.

biobank uk

+ Our Future Health

RECOVERY
Randomised Evaluation of COVID-19 Therapy

Accessing, linking and analysing more granular and detailed data: example of Scottish medical imaging archive

Radiology: Artificial Intelligence

DATA RESOURCES

The Scottish Medical Imaging Archive: 57.3 Million Radiology Studies Linked to Their Medical Records

Radiology: Artificial Intelligence 2024; 6(1):e220266

Dementia

Scientists to use AI to analyse 1.6m brain scans to develop tool predicting dementia risk

Researchers will use artificial intelligence to match image data of patients from Scotland with linked health records

Retinal imaging data to predict neurodegenerative diseases

BMJ Open

AlzEye: longitudinal record-level linkage of ophthalmic imaging and hospital admissions of 353 157 patients in London, UK

Siegfried Karl Wagner [1,2] Fintan Hughes,[3] Mario Cortina-Borja,[4] Nikolas Pontikos [1,2] Robbert Struyven,[1,2] Xiaoxuan Liu [5,6,7] Hugh Montgomery,[8] Daniel C Alexander [9] Eric Topol [10] Steffen Erhard Petersen [11,12] Konstantinos Balaskas [1,2,13] Jack Hindley,[14] Axel Petzold [1,15,16] Jugnoo S Rahi [1,2,17,18,19] Alastair K Denniston,[5,6,7] Pearse A Keane [1,2,13]

INSIGHT
Discovery | Innovation | Care

Neurology®

August 21, 2023 RESEARCH ARTICLE
OPEN ACCESS

Retinal Optical Coherence Tomography Features Associated With Incident and Prevalent Parkinson Disease

Siegfried Karl Wagner, David Romero-Bascones, Mario Cortina-Borja, Dominic J Williamson, Robbert R Struyven, Yukun Zhou, Salil Patel, Rimona S Weil, Chrystalina A Antoniades, Eric J Topol, Edward Korot, Paul J Foster, Konstantinos Balaskas, Unai Ayala, Maitane Barrenechea, Iñigo Gabilondo, Anthony HV Schapira, Anthony P Khawaja, Praveen J Patel, Jugnoo S Rahi, Alastair K Denniston, Axel Petzold, Pearse Andrew Keane, for UK Biobank Eye & Vision Consortium

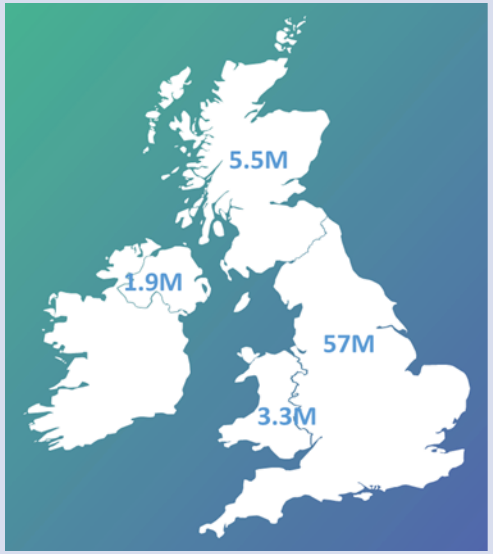
First published August 21, 2023, DOI: https://doi.org/10.1212/WNL.0000000000207727

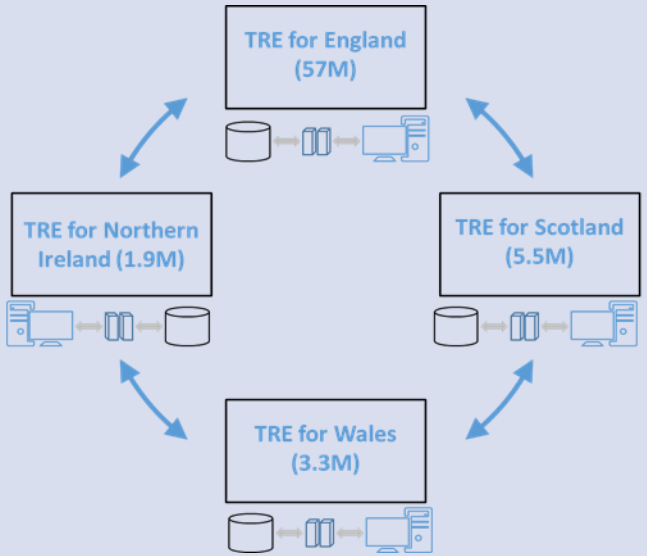
Study shows eye scans could diagnose Parkinson's disease early by seven years

HEALTH | SCIENCE | PARKINSON'S DISEASE | Monday 21 August 2023 at 11:12pm

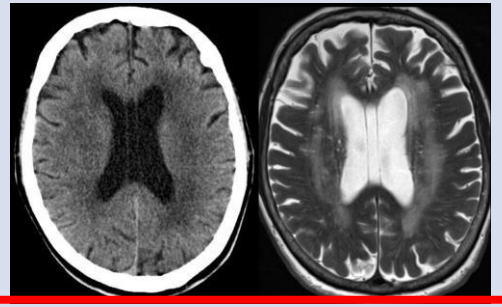
Health-relevant data across the UK

NHS: NATIONALLY COLLATED, POPULATION-WIDE DATA





NHS: LOCAL & REGIONAL DATA



NON-NHS: HEALTH-RELEVANT ADMINISTRATIVE DATA

Census, education, disability, income….

ADR UK

Research Data Scotland

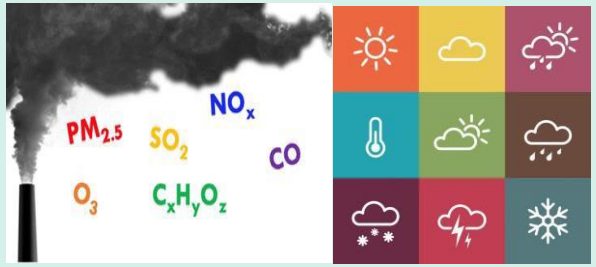
SAIL DATABANK

Northern Ireland Statistics & Research Agency

Office for National Statistics

PERSONAL MONITORING DATA

ENVIRONMENTAL DATA



RESEARCH GENERATED DATA

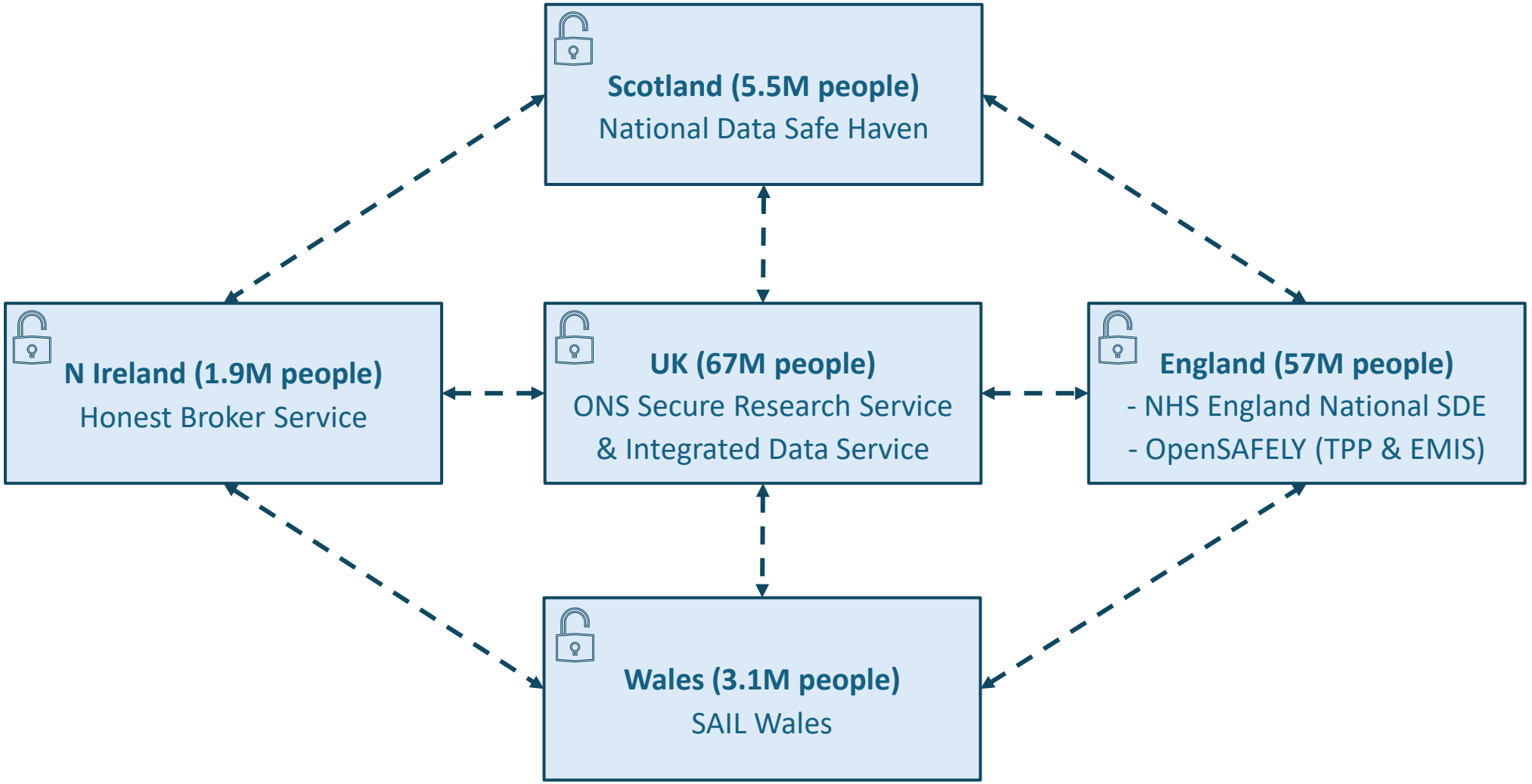
Cohorts, biobanks, clinical trials, e.g.

biobank uk

\+ Our Future Health

RECOVERY
Randomised Evaluation of COVID-19 Therapy

Linking whole population healthcare to other administrative data sources to understand the wider determinants of health



Linking data across sectors to understand the wider determinants of health

Association of school absence and exclusion with recorded neurodevelopmental disorders, mental disorders, or self-harm: a nationwide, retrospective, electronic cohort study of children and young people in Wales, UK

Ann John, Yasmin Friedmann, Marcos DelPozo-Banos, Aura Frizzati, Tamsin Ford, Anita Thapar

Lancet Psychiatry 2022;
9: 23-34

International Journal of Epidemiology, 2022, 17–17f
doi: 10.1093/ije/dyab149
Advance Access Publication Date: 11 November 2021
Data Resource Profile

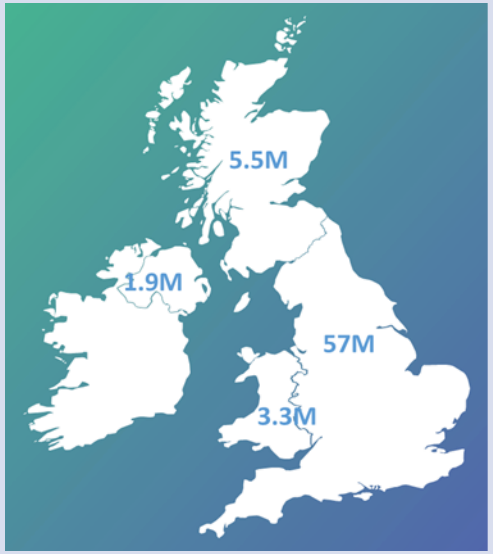
Data Resource Profile

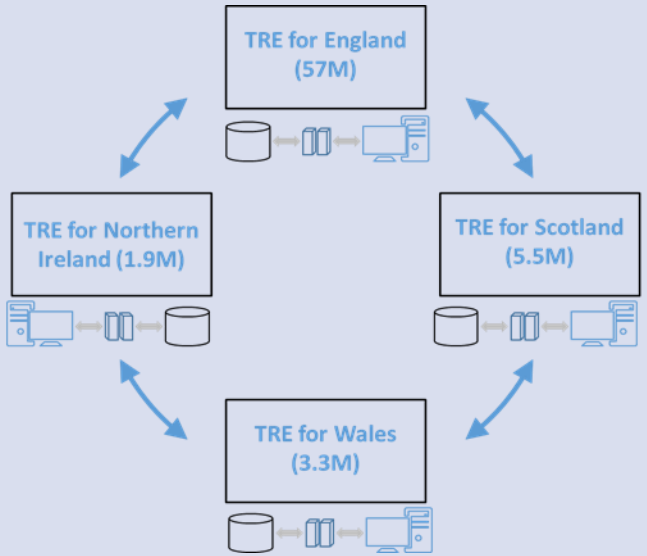
Data Resource Profile: The Education and Child Health Insights from Linked Data (ECHILD) Database

Louise Mc Grath-Lone ,[1]* Nicolás Libuy,[1,2] Katie Harron ,[3] Matthew A Jay ,[3] Linda Wijlaars,[3] David Etoori,[1] Matthew Lilliman,[3] Ruth Gilbert,[3] and Ruth Blackburn[1]

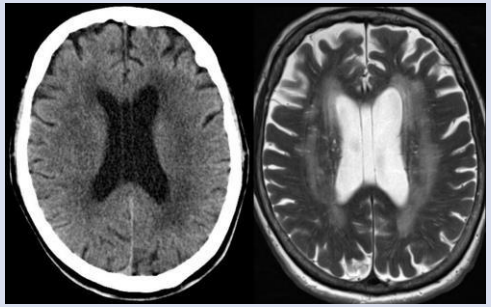
Health-relevant data across the UK

NHS: NATIONALLY COLLATED, POPULATION-WIDE DATA





NHS: LOCAL & REGIONAL DATA



NON-NHS: HEALTH-RELEVANT ADMINISTRATIVE DATA

Census, education, disability, income….

ADR UK

Research Data Scotland

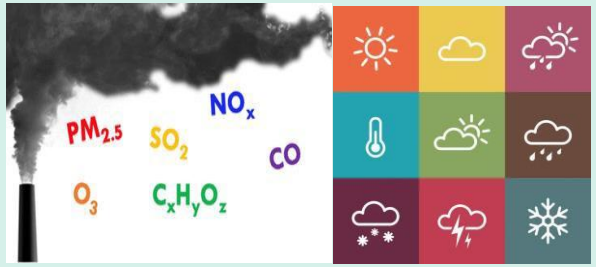
SAIL DATABANK

Northern Ireland Statistics & Research Agency

Office for National Statistics

PERSONAL MONITORING DATA

ENVIRONMENTAL DATA



RESEARCH GENERATED DATA

Cohorts, biobanks, clinical trials, e.g.

biobank uk

\+ Our Future Health

RECOVERY
Randomised Evaluation of COVID-19 Therapy

Routinely collected data to enhance large-scale clinical trials

RECOVERY

Centralised services providing national linked data on hospitalisations and deaths to ascertain outcomes in 50,000 participants

NHS Galleri Trial

Centralised service using national data to invite millions of people and recruit 140,000 participants to a multi-cancer screening blood test trial

Large population-based cohort studies with detailed multi-modal data, samples, and long-term follow-up of participants using linked routine health data

biobank uk

500,000
40-69 years

generation scotland

35,000 → 40,000
12+ years

\+
Our Future Health

1.5 → 5 million
18-70 years

UK Research and Innovation
Adolescent Health Study

0 → 100,000+
8-18 years

Advantages of PROSPECTIVE population-based cohorts for studying the determinants of disease

Risk factors can be measured long before disease develops, avoiding “reverse causation” and recall bias (especially for diseases that affect cognition / recall)

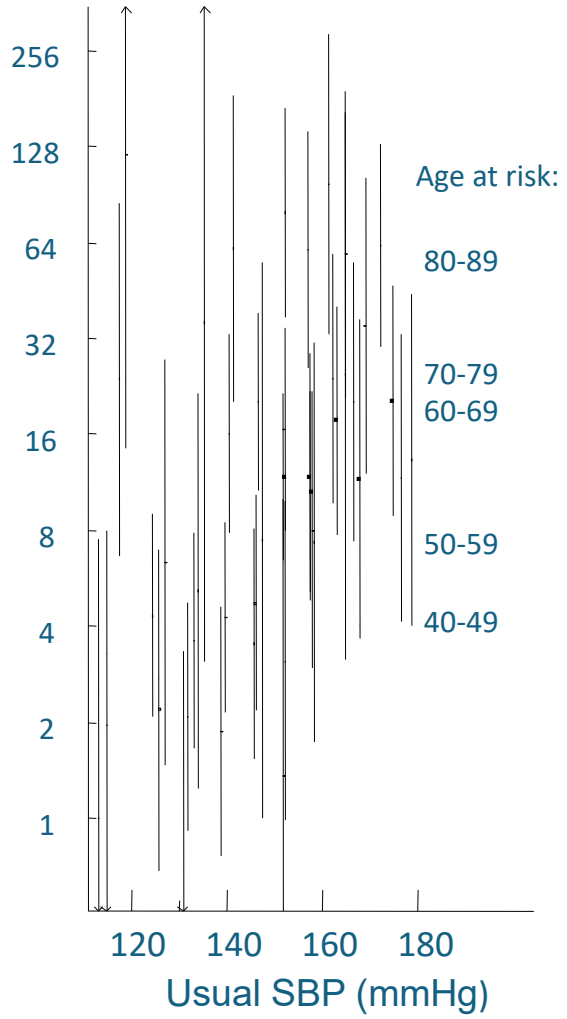
Non-disease controls come from the same population as the disease cases

Can assess the effects of any potential risk factor or cause (e.g. smoking, the pill, a genetic variant) on many different diseases (e.g. lung disease, cancer, vascular disease, dementia)

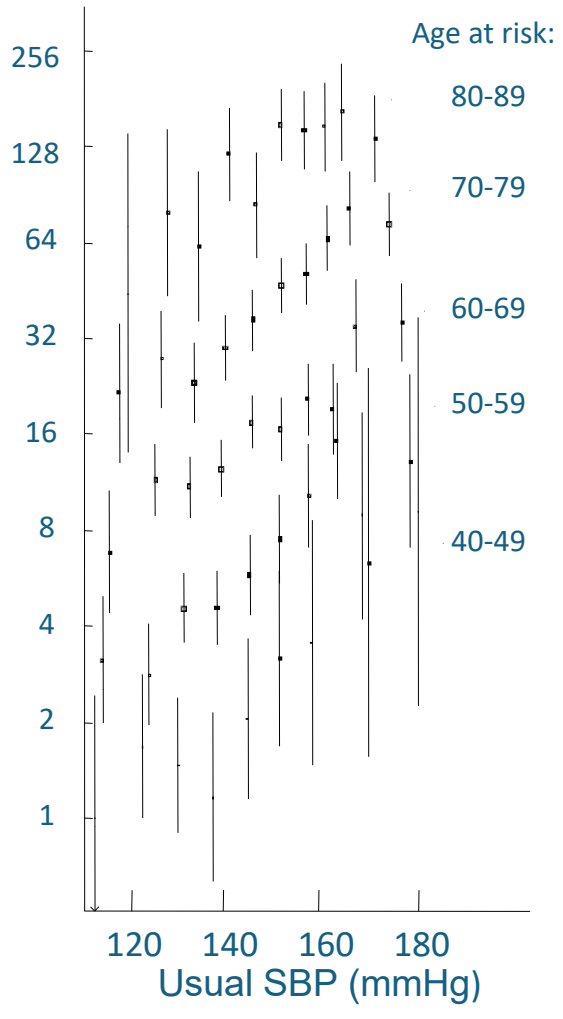
But prospective cohorts need to be LARGE since only a proportion of the participants develop any particular disease during prolonged follow-up

Prospective studies need to be very large: example of blood pressure and risk of coronary heart disease

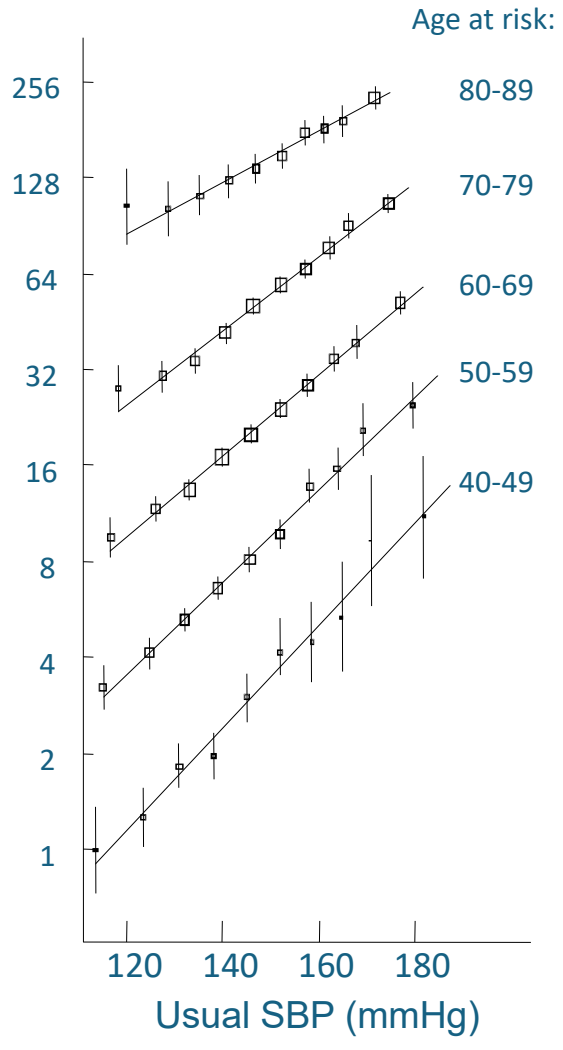
5000 people



50,000 people



500,000 people

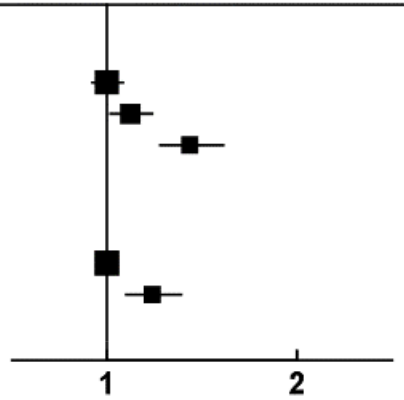


Prospective studies need to be very large and with long follow-up

Risk factors and dementia risk:
1.3 million middle aged women followed for 16 years
with 24,000 cases of incident (new onset) dementia

< 10 YEARS AFTER BASELINE
(n=4537)

Exposure	Number with dementia	Relative risk (99% CI)
Smoking		
never	1965	1.00 (0.92, 1.08)
past	1228	1.12 (1.02, 1.24)
current	996	1.44 (1.28, 1.61)
Physical activity		
active	2622	1.00
inactive	1433	1.24 (1.10, 1.39)

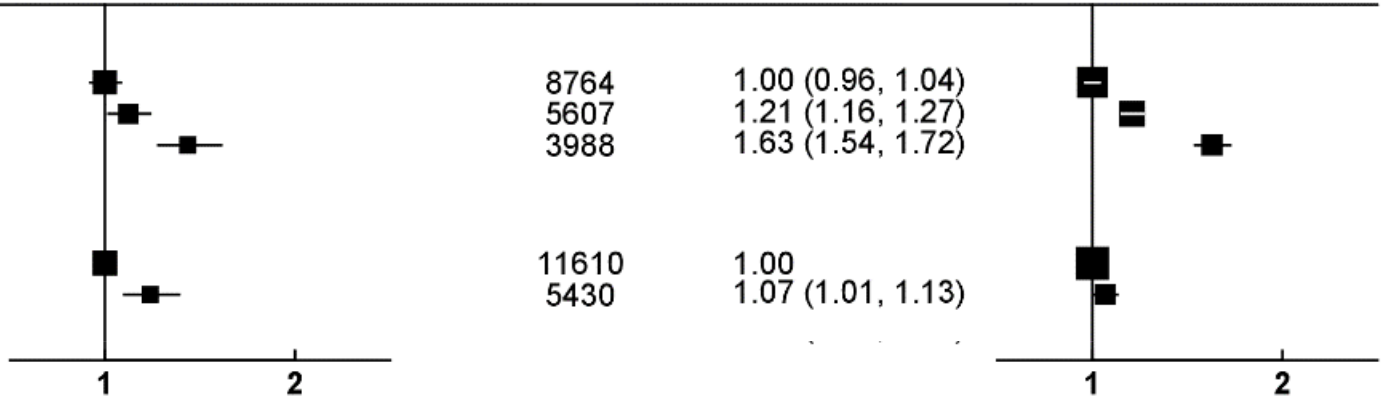


Million Women Study, Neurology 2019

Prospective studies need to be very large and with long follow-up

Risk factors and dementia risk:
1.3 million middle aged women followed for 16 years
with 24,000 cases of incident (new onset) dementia

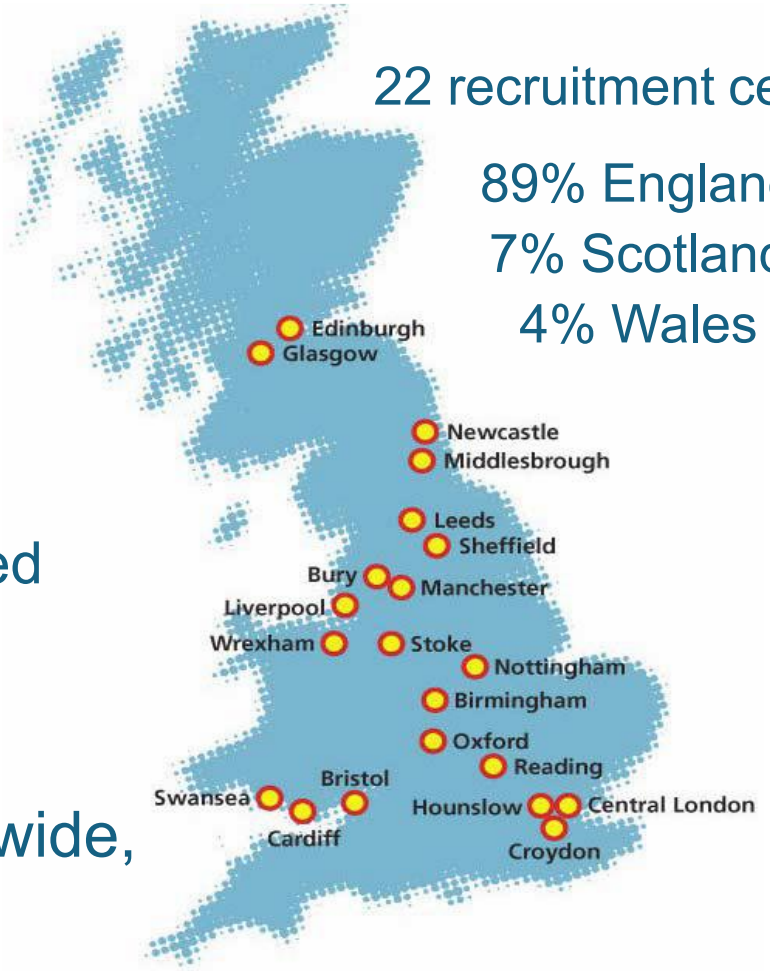
Exposure	< 10 YEARS AFTER BASELINE (n=4537) Number with dementia	Relative risk (99% CI)	≥ 10 YEARS AFTER BASELINE (n=19803) Number with dementia	Relative risk (99% CI)
Smoking				
never	1965	1.00 (0.92, 1.08)	8764	1.00 (0.96, 1.04)
past	1228	1.12 (1.02, 1.24)	5607	1.21 (1.16, 1.27)
current	996	1.44 (1.28, 1.61)	3988	1.63 (1.54, 1.72)
Physical activity				
active	2622	1.00	11610	1.00
inactive	1433	1.24 (1.10, 1.39)	5430	1.07 (1.01, 1.13)



Million Women Study, Neurology 2019

UK Biobank:
an open access prospective study with large size and extensive breadth and depth of data

- Very large population-based cohort
  - 503,000 participants
- Extensive and detailed questions, measures and samples, multimodal imaging, wearables
- Follow-up of participants' health status
  - comprehensive linkage to wide range of health-related records through national electronic systems
  - online questionnaires
- Open access resource (>30,000 researchers worldwide, 1000s of publications)
  - see http://www.ukbiobank.ac.uk



22 recruitment centres

89% England
7% Scotland
4% Wales

biobank uk

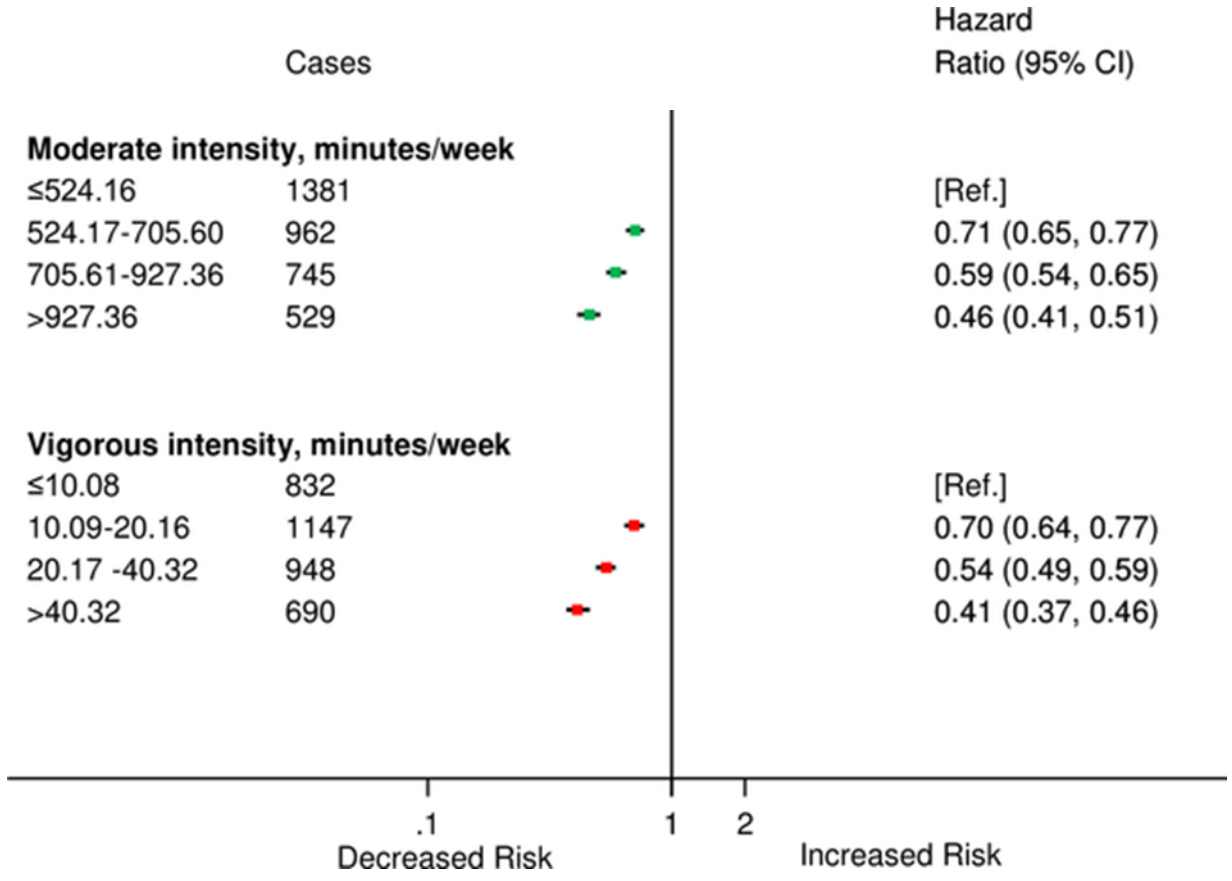
Linked routine health data to follow the health of participants in large population-based longitudinal research studies: impact of physical activity on health in UK Biobank

RESEARCH ARTICLE

Accelerometer measured physical activity and the incidence of cardiovascular disease: Evidence from the UK Biobank cohort study

Rema Ramakrishnan[1,2], Aiden Doherty[3,4,5], Karl Smith-Byrne[6], Kazem Rahimi[1,5,7,8], Derrick Bennett[5,9], Mark Woodward[10,11,12], Rosemary Walmsley[3,4], Terence Dwyer[1,13]*

- Compared with the least active people, the most active people had less than half the risk of stroke and heart attack.
- Protective effect of physical activity is much stronger than previously thought.



biobank uk

These many sources of data can be used to save and improve people’ lives

There are great examples that demonstrate what’s possible

But they are far too often exceptions rather than business as usual

Still great complexity, friction, delays and gaps…

‘Sudlow Review’: Five recommendations for system-wide reform

1) Coordinated strategy for health data as a critical national infrastructure
2) A national health data service with senior accountable leadership, to include
    - Single national health data access system
    - Streamlined and standardised data governance and access
    - National system for general practice data for full range of beneficial uses
    - Better, faster access to other major national and regional NHS data assets
    - National scale linkage to health-relevant data from beyond the health and care system
3) Coordinated engagement with patients, public, health professionals, policy makers and politicians
4) UK-wide approach for data access processes and proportionate data governance
5) UK-wide system for standards and accreditation of secure data environments

Influencing UK Government policy

- Review received positively across NHS, industry, academia, charities, public and patient groups, and relevant government departments
- Aligns with government's economic growth agenda and with proposed shifts in healthcare (hospital to community, analogue to digital, treatment to prevention)
- Has informed Life Sciences Sector Plan and 10 Year Health Plan for the NHS

Influencing UK Government policy

medConfidential
@medConfidential

<10 minutes after publication:

The drivel of the HDR Sudlow review is out

< 10 hours after publication:

The Sudlow review in short: give more data to biobank so they can give it to racists

Influencing investment in delivery

April 2025 - Prime Minister announced £600 million for a national Health Data Research Service

GOV.UK

Press release

Prime Minister turbocharges medical research

Better and faster access to NHS data for researchers, with gold standard security and privacy measures.

HDRUK
Health Data Research UK

A defining moment: Government announces a Health Data Research Service

7 April 2025

Wellcome is partnering with the UK Government to establish a new £600m Health Data Research Service; an exciting development following a central recommendation of the Sudlow Review.

wellcome

National data service will simplify access to health data for research

Wellcome is partnering with the UK Government to establish a new up to £600 million health data research service. It will simplify secure access to health data and speed up research to better understand ill health, opening opportunities to develop prevention, diagnosis and treatment tools for diseases.

Key features of proposed national Health Data Research Service

- Government-owned company
- UK-wide remit
- Core, must-have, whole population, national datasets:
  - general practice
  - hospital episodes
  - medicines/devices
  - deaths
- Underpinned by:
  - principle of public data for public benefit
  - secure data environments to ensure privacy and security of individuals' data
  - supportive legislation and streamlined regulation
  - ongoing public/patient/professional consultation, engagement and involvement