Health Data Research UK and the British Heart Foundation Data Science Centre – Phenotype Library

Angela M Wood

13th Nov 2024

ESC
European Society of Cardiology

British Heart Foundation Data Science Centre

Led by Health Data Research UK

ESC

Cardiovascular Round Table

- UK centre of expertise for cardiovascular data science
- Mission: harnessing the power of health data to improve heart and circulatory health.

Whole Population Data

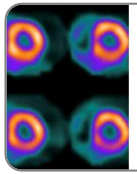
Better use of nationally-collated, structured, coded data: accessing, improving and using linked, national,…

Read more

Defining Disease

Developing methods to define cardiovascular health and disease in computable form through a collaborative network of…

Read more



Imaging

Better use of unstructured data: addressing the challenges of accessing, improving and using unstructured data, for…

Read more

Cohorts

Facilitating the linkage of large, 'omics-rich' cohorts to electronic health records to better understand the causes of…

Read more

Clinical Trials

Developing platforms for efficient, cost-effective trials, using routine health data to recruit and follow patients with…

Read more

Smartphones and Wearables

Exploring how data from apps and wearables, linked to other health datasets, can inform trajectories of cardiovascular…

Read more

Diabetes Data Science Catalyst

This exciting partnership between the BHF Data Science Centre, Diabetes UK and HDR UK aims to develop improvements in…

Read more

Stroke Data Science Catalyst

This five-year partnership between the British Heart Foundation (BHF) Data Science Centre, Health Data Research UK (HDR…

Read more

Kidney Data Science Catalyst

This partnership between the BHF Data Science Centre, Kidney Research UK and HDR UK will enable researchers to securely…

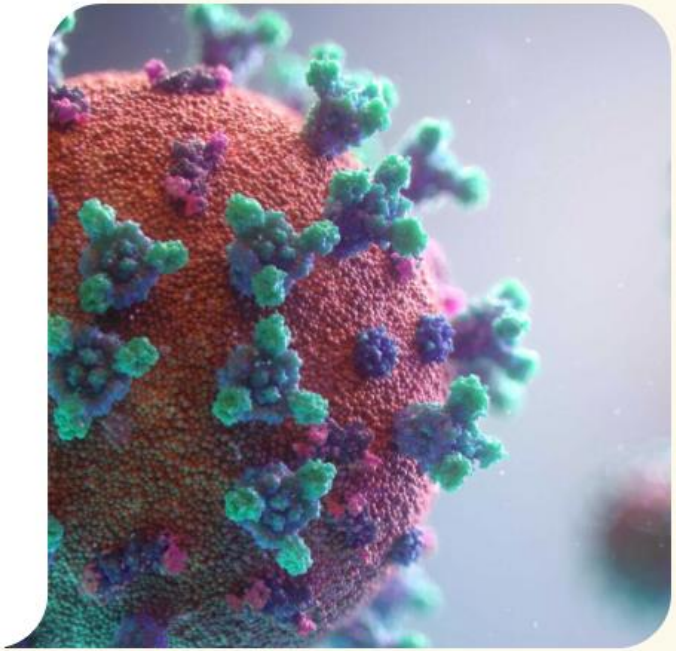
Read more

Whole Population Data

Areas of work

CVD-COVID-UK / COVID-IMPACT

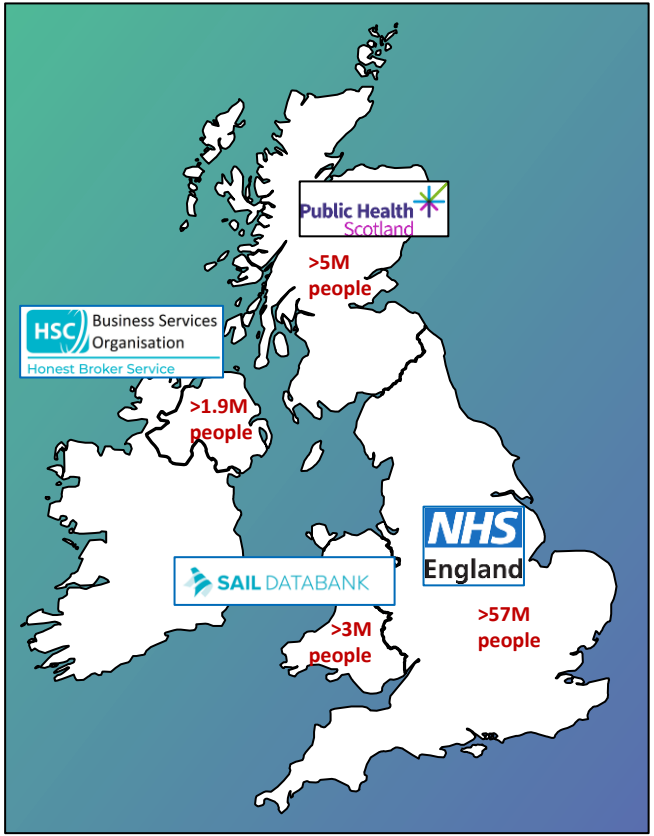
CVD-COVID-UK aims to understand the relationship between COVID-19 and cardiovascular diseases such as heart attack, heart failure, stroke, and blood clots in the lungs through analyses of de-identified, linked, nationally collated healthcare datasets across the four nations of the UK. COVID-IMPACT is an expansion of this approach to address research questions looking at the impact of COVID-19 on other health conditions and their related risk factors.



Uniting the nations' electronic health records for research

ESC Cardiovascular Round Table

- Secure access to multiple linked health datasets
  - Primary care
  - Hospital
  - Deaths
  - Medications
  - Specialist registries / audits
  - COVID-19 vaccination and lab test data
  - Climate/pollution data (linked on 1km^2 areas)
- Enables whole population research on **population of >67 million people**:
  Statistically powerful
  - Comprehensive information on characteristics and health outcomes
  - Includes all ages groups, ethnicities, geographic locations, socioeconomic, health and personal characteristics
- Datasets updated regularly



Research Activity

90+ projects

80+ approved researchers

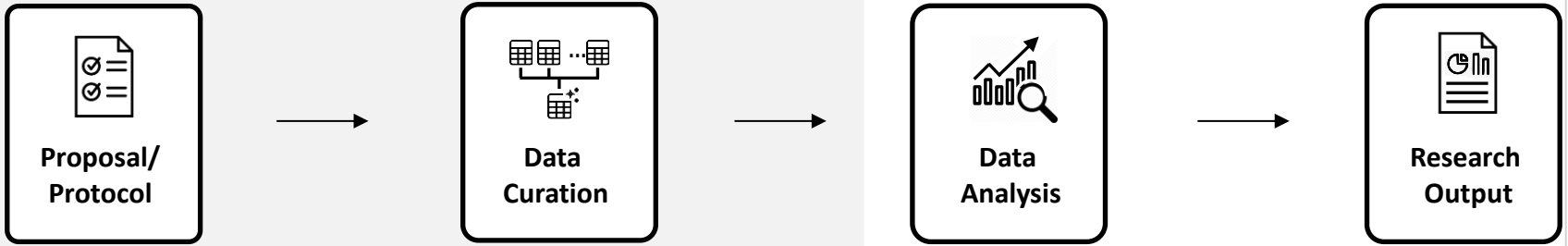
50+ NHS and academic organisations

40 publications and preprints

Multiple policies

Project Journey

ESC
Cardiovascular Round Table

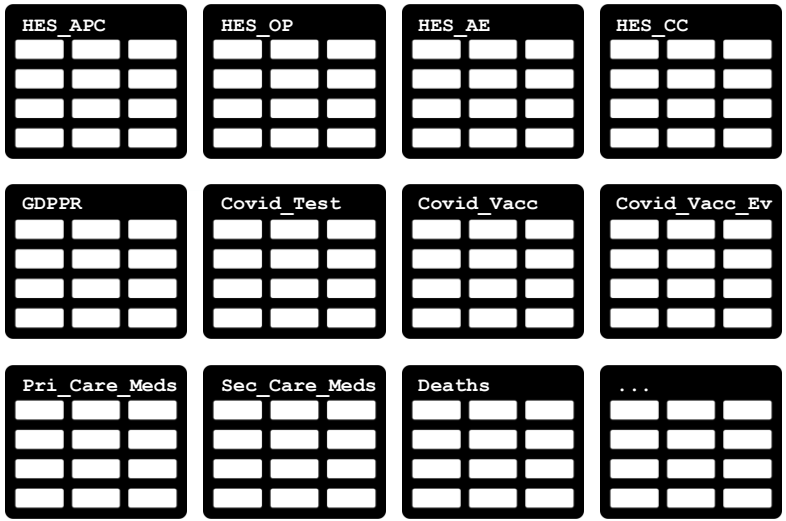


"It has been estimated that **80%** of the work for data science with NHS records is spent on **data preparation**."

Goldacre, B & Morley, J (2022). Better, Broader, Safer: Using health data for research and analysis.

Data curation pipeline

Provisioned data

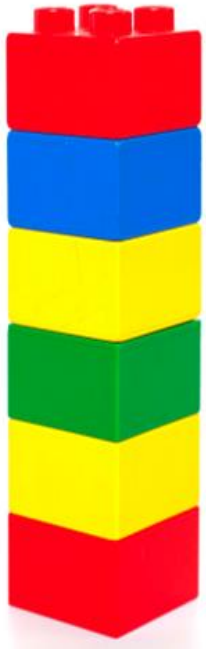


Data curation pipeline

- Data exploration
- Data wrangling
- Data cleaning
- Data harmonisation
- Data phenotyping
- Data checks/validation
- Data visualisation

Analysis-ready data

HDR UK Phenotype Library: systematic curation of EHR phenotyping algorithms



- Primary care clinical events (SNOMED, Read)
- Cause of death (ICD-10)
- In-patient hospital diagnoses (ICD-10)

Dr Jackie MacArthur

Prof Spiros Denaxas

https://phenotypes.healthdatagateway.org/phenotypes/PH942/version/2120/detail/

Linking data from different healthcare settings to ascertain incident cardiovascular events

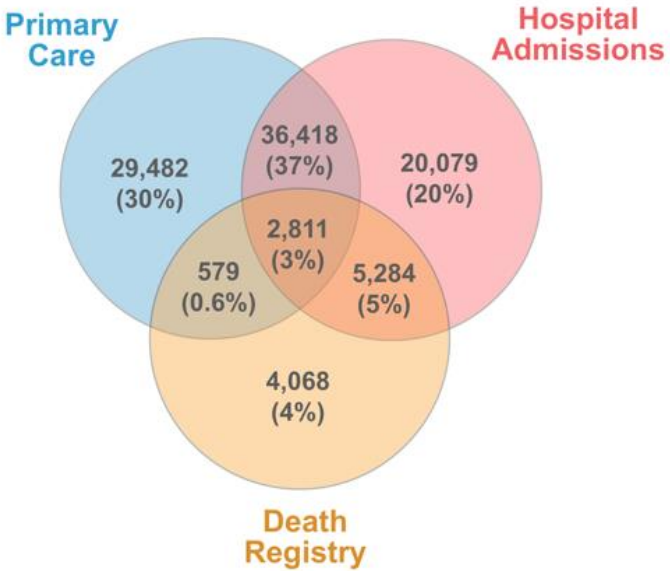
British Heart Foundation Data Science Centre
Led by Health Data Research UK

HDRUK Health Data Research UK

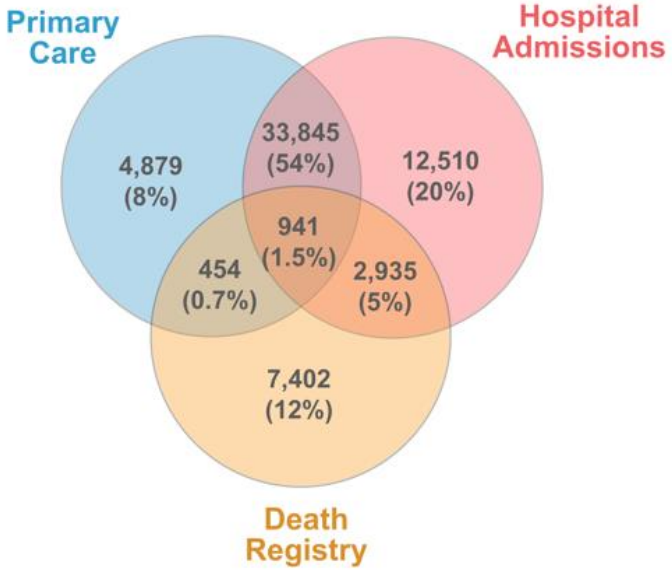
ESC Cardiovascular Round Table

January to October 2020

Incident Stroke / TIA (n=98,721)



Incident myocardial infarction (n=63,966)



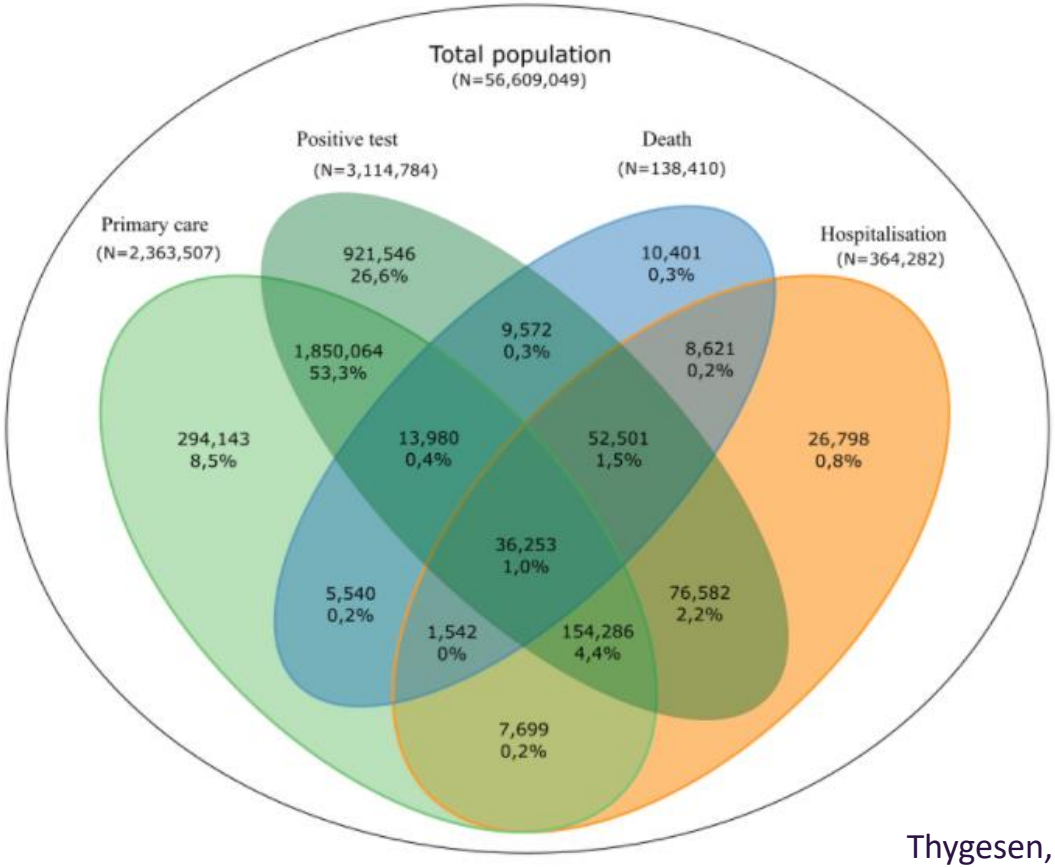
BMJ 2021; 373 doi: https://doi.org/10.1136/bmj.n826

Linking data from different healthcare settings to ascertain COVID-19 infections at population scale

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

HDRUK Health Data Research UK

ESC Cardiovascular Round Table



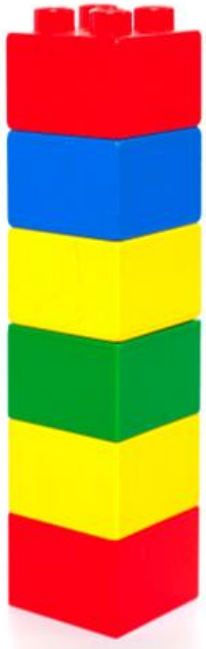
3.5 million people with COVID-19 by mid-Feb 2021:

- 3.1 million with a positive test
- 2.4 million diagnosed in primary care
- 364,000 hospitalised

Thygesen, Tomlinson & Denaxas, *The Lancet Digital Health*, 2022

HDR UK Phenotype Library: systematic curation of EHR phenotyping algorithms

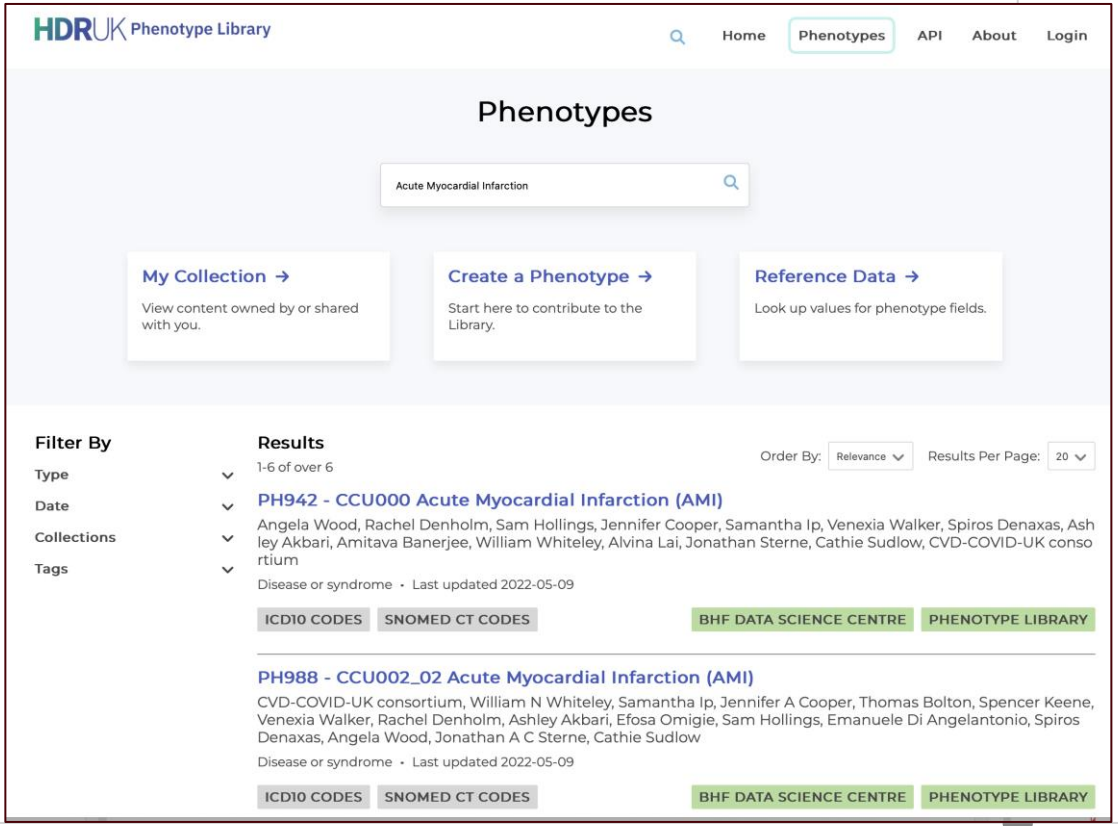
ESC
Cardiovascular Round Table



→ Primary care clinical events (SNOMED, Read)

→ Cause of death (ICD-10)

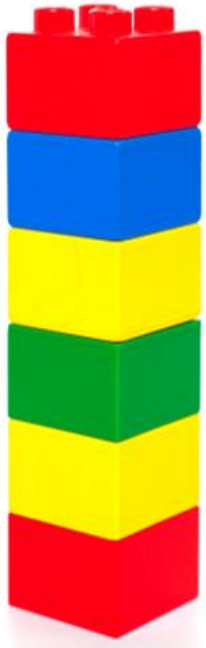
→ In-patient hospital diagnoses (ICD-10)



https://phenotypes.healthdatagateway.org/phenotypes/PH942/version/2120/detail/

HDR UK Phenotype Library: systematic curation of EHR phenotyping algorithms

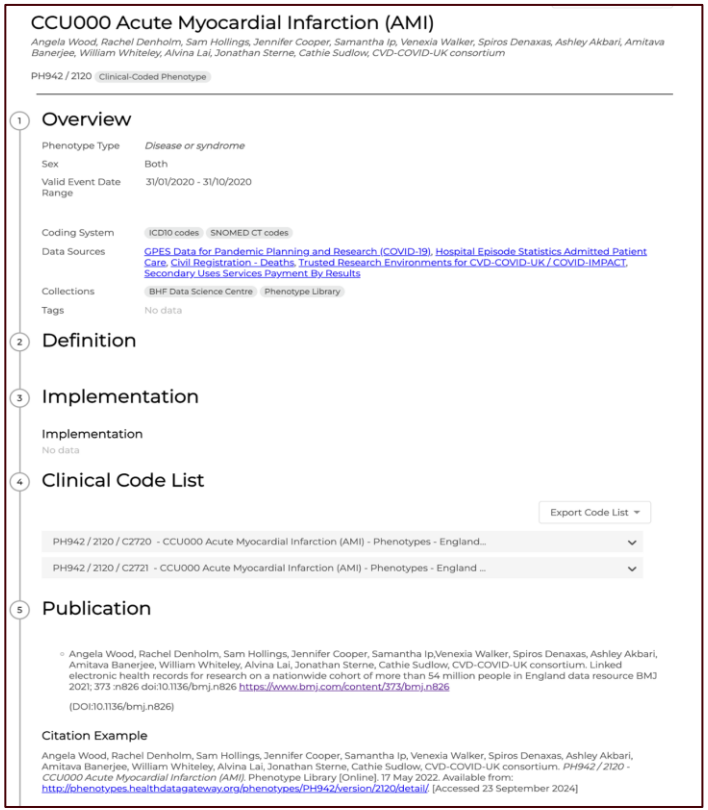
ESC
Cardiovascular Round Table



→ Primary care clinical events (SNOMED, Read)

→ Cause of death (ICD-10)

→ In-patient hospital diagnoses (ICD-10)



CCU000 Acute Myocardial Infarction (AMI)

Angela Wood, Rachel Denholm, Sam Hollings, Jennifer Cooper, Samantha Ip, Venexia Walker, Spiros Denaxas, Ashley Akbari, Amitava Banerjee, William Whiteley, Alvina Lai, Jonathan Sterne, Cathie Sudlow, CVD-COVID-UK consortium

PH942 / 2120 Clinical-Coded Phenotype

1 Overview

Phenotype Type	*Disease or syndrome*
Sex	Both
Valid Event Date Range	31/01/2020 - 31/10/2020
Coding System	ICD10 codes SNOMED CT codes
Data Sources	GPES Data for Pandemic Planning and Research (COVID-19), Hospital Episode Statistics Admitted Patient Care, Civil Registration - Deaths, Trusted Research Environments for CVD-COVID-UK / COVID-IMPACT, Secondary Uses Services Payment By Results
Collections	BHF Data Science Centre Phenotype Library
Tags	No data

2 Definition

3 Implementation

Implementation
No data

4 Clinical Code List

Export Code List

PH942 / 2120 / C2720 - CCU000 Acute Myocardial Infarction (AMI) - Phenotypes - England...

PH942 / 2120 / C2721 - CCU000 Acute Myocardial Infarction (AMI) - Phenotypes - England ...

5 Publication

- Angela Wood, Rachel Denholm, Sam Hollings, Jennifer Cooper, Samantha Ip,Venexia Walker, Spiros Denaxas, Ashley Akbari, Amitava Banerjee, William Whiteley, Alvina Lai, Jonathan Sterne, Cathie Sudlow, CVD-COVID-UK consortium. Linked electronic health records for research on a nationwide cohort of more than 54 million people in England data resource BMJ 2021; 373 :n826 doi:10.1136/bmj.n826 https://www.bmj.com/content/373/bmj.n826

(DOI:10.1136/bmj.n826)

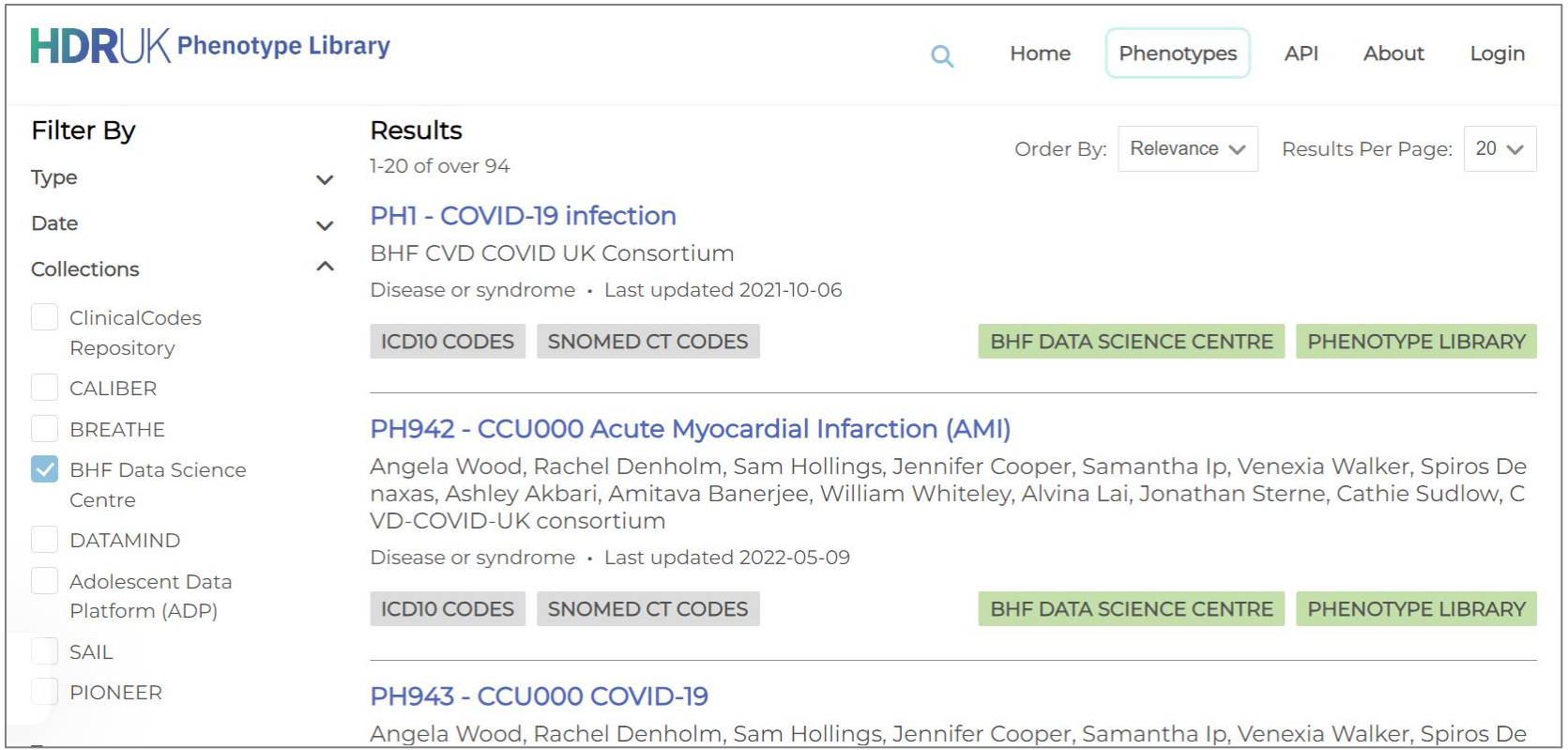
Citation Example

Angela Wood, Rachel Denholm, Sam Hollings, Jennifer Cooper, Samantha Ip, Venexia Walker, Spiros Denaxas, Ashley Akbari, Amitava Banerjee, William Whiteley, Alvina Lai, Jonathan Sterne, Cathie Sudlow, CVD-COVID-UK consortium. *PH942 / 2120 - CCU000 Acute Myocardial Infarction (AMI).* Phenotype Library [Online]. 17 May 2022. Available from: http://phenotypes.healthdatagateway.org/phenotypes/PH942/version/2120/detail/ [Accessed 23 September 2024]

https://phenotypes.healthdatagateway.org/phenotypes/PH942/version/2120/detail/

BHF DSC collection in Phenotype Library

ESC Cardiovascular Round Table



Prevalence of codes in SDE

Select Coding System: ICD10

Overall Count: Records

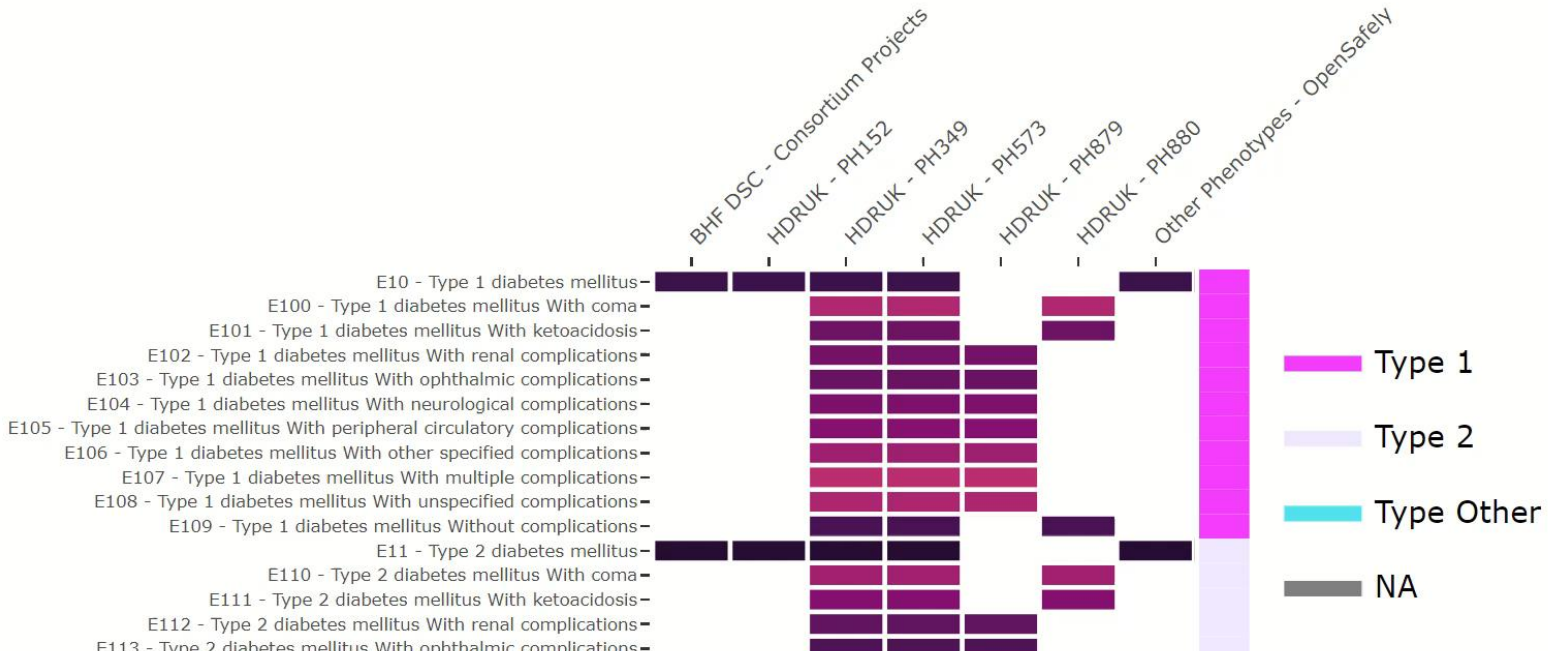
Count Threshold: 10

Order By: Code

Phenotypes: Select

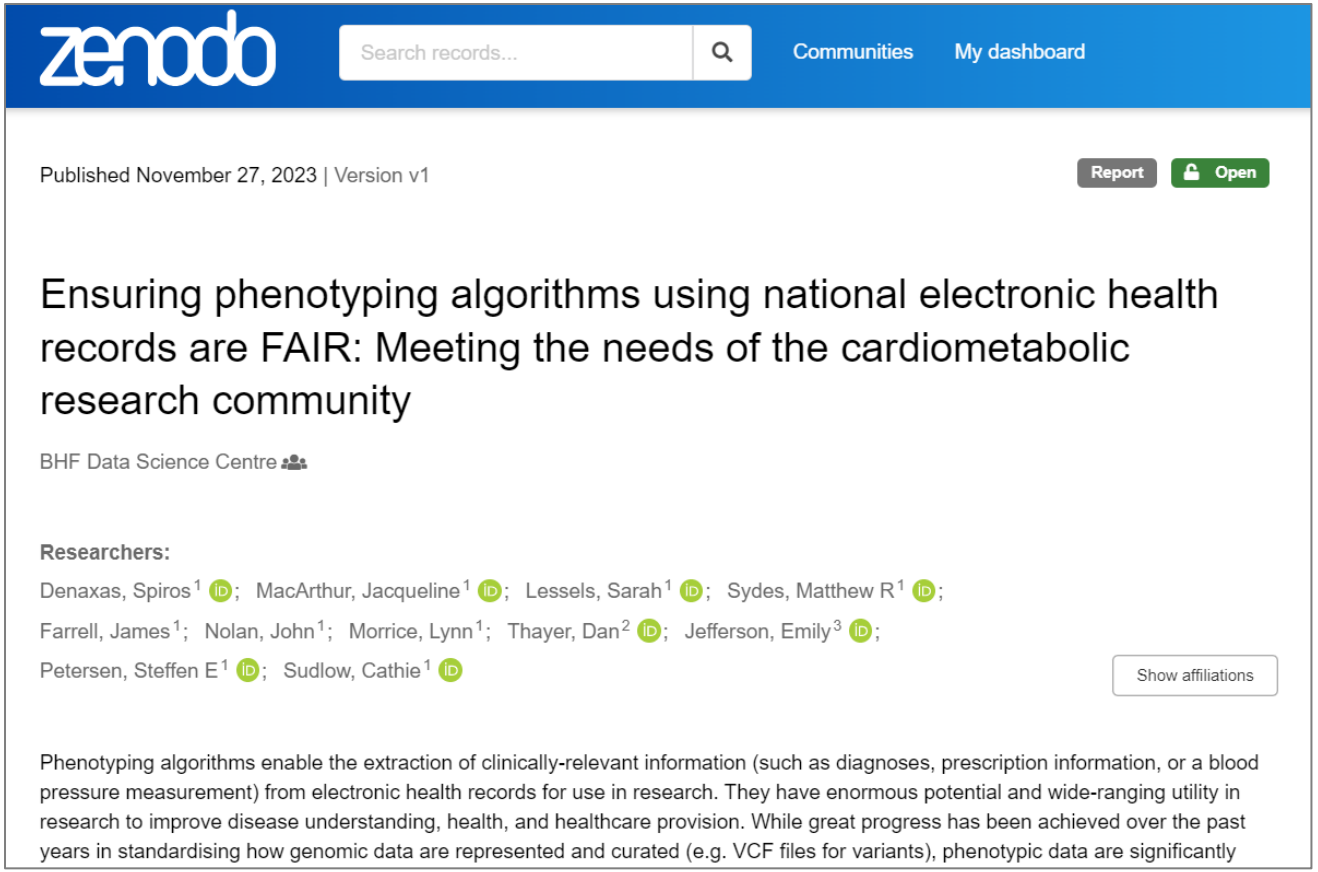
Refset: In × Not in ×

Diabetes Type:



Ensuring phenotyping algorithms are FAIR

ESC Cardiovascular Round Table



zenodo

Search records...

Communities My dashboard

Published November 27, 2023 | Version v1

Report Open

Ensuring phenotyping algorithms using national electronic health records are FAIR: Meeting the needs of the cardiometabolic research community

BHF Data Science Centre

Researchers:

Denaxas, Spiros[1]; MacArthur, Jacqueline[1]; Lessels, Sarah[1]; Sydes, Matthew R[1]; Farrell, James[1]; Nolan, John[1]; Morrice, Lynn[1]; Thayer, Dan[2]; Jefferson, Emily[3]; Petersen, Steffen E[1]; Sudlow, Cathie[1]

Show affiliations

Phenotyping algorithms enable the extraction of clinically-relevant information (such as diagnoses, prescription information, or a blood pressure measurement) from electronic health records for use in research. They have enormous potential and wide-ranging utility in research to improve disease understanding, health, and healthcare provision. While great progress has been achieved over the past years in standardising how genomic data are represented and curated (e.g. VCF files for variants), phenotypic data are significantly

Standards for phenotyping algorithms

Recommended dimensions:

1. Complexity
2. Performance
3. Efficiency
4. Implementability
5. Maintenance

JOURNAL ARTICLE

Improving reporting standards for phenotyping algorithm in biomedical research: 5 fundamental dimensions

Get access

Wei-Qi Wei, MD, PhD, Robb Rowley, MD, Angela Wood, PhD, Jacqueline MacArthur, PhD, Peter J Embi, MD, Spiros Denaxas, PhD

Journal of the American Medical Informatics Association, Volume 31, Issue 4, April 2024, Pages 1036–1041, https://doi.org/10.1093/jamia/ocae005
Published: 24 January 2024 **Article history**

Cite Permissions Share

Abstract

Introduction

Phenotyping algorithms enable the interpretation of complex health data and definition of clinically relevant phenotypes; they have become crucial in biomedical research. However, the lack of standardization and transparency inhibits the cross-comparison of findings among different studies, limits large scale meta-analyses, confuses the research community, and prevents the reuse of algorithms, which results in duplication of efforts and the waste of valuable resources.

Burden of cardiovascular diseases in England

ESC
Cardiovascular Round Table

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

About | **Compare Diseases** | Explore a Single Disease | Methodology

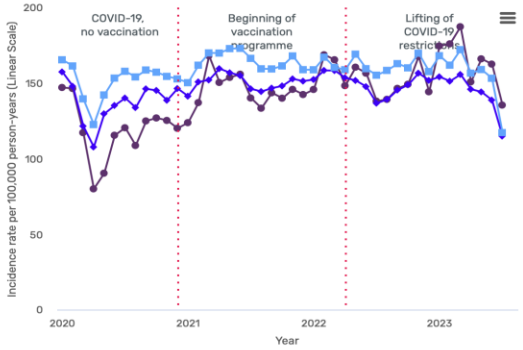
Date Range: 2020 – 2023

Select Diseases:
- Heart failure NOS ×
- Myocardial infarction ×
- Ischaemic stroke syndrome ×

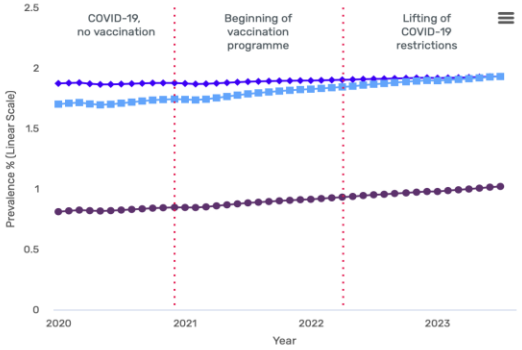
- Show COVID-19 timeline
- Disease labels
- Log scale
- Show all trajectories
- Show trend
- 6-month moving average

Heart failure NOS
Myocardial infarction
Ischaemic stroke syndrome

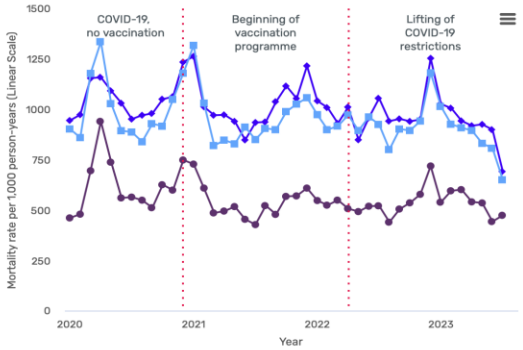
Incidence



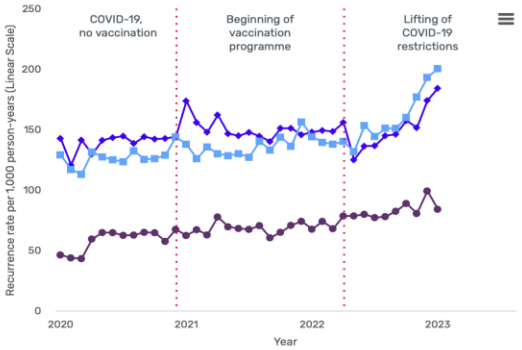
Prevalence



Case Fatality



Recurrence



Prof William Whiteley

Burden of cardiovascular diseases in England

ESC
Cardiovascular Round Table

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

About | Compare Diseases | **Explore a Single Disease** | Methodology

Date Range: 2020 – 2023

Select Disease: Heart failure NOS

Select Characteristic: Ethnicity

Show COVID-19 timeline

Characteristic labels

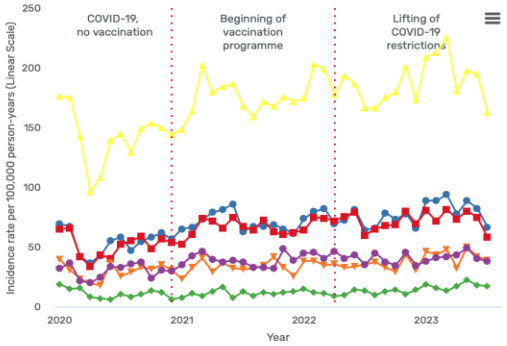
Log scale

Show trend

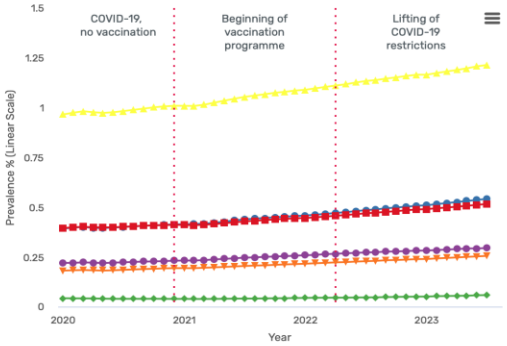
6-month moving average

Ethnicity
Asian or Asian British | Missing | Other
Black or Black British | Mixed | White

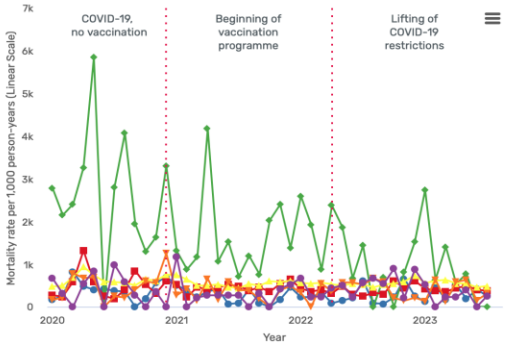
Incidence



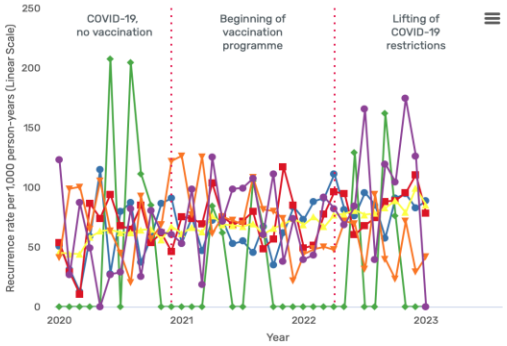
Prevalence



Case Fatality



Recurrence



Summary

- HDR UK Phenotype Library: systematic curation of EHR phenotyping algorithms
- Phenotyping algorithms need to be Findable, Accesible, Interoperable, and Reusable (FAIR) and adhere to newly recommended standards

- **Builds an essential and high-quality foundation for all subsequent analyses**

Thank you for listening

www.bhfdatasciencecentre.org
bhfdsc@hdruk.ac.uk
@BHFDataScience