



**British Heart Foundation  
Data Science Centre**

Led by Health Data Research UK

# Mapping the population data of England into OMOP CDM

Leveraging big data technologies in a  
Secure Data Environment (SDE)

Mehrdad Mizani

21/11/2024

**HDRUK**  
Health Data Research UK



[www.bhfdatasciencecentre.org](http://www.bhfdatasciencecentre.org)

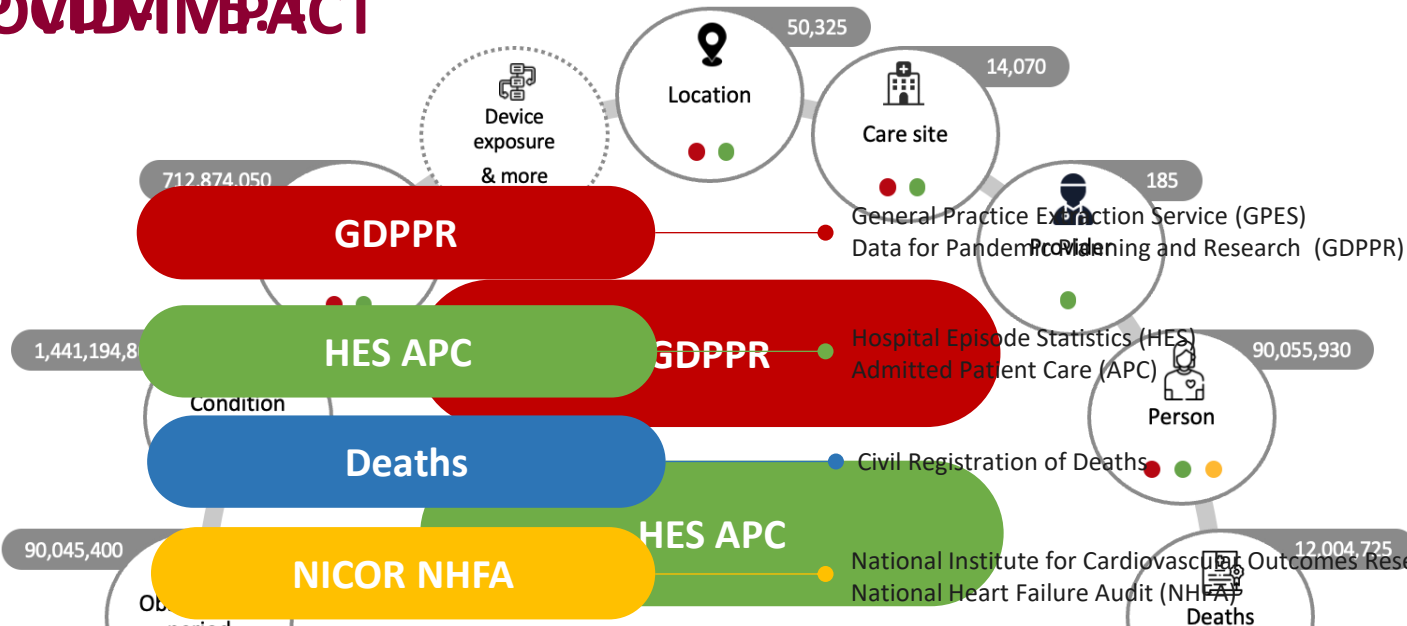


[bhfdsc@hdruk.ac.uk](mailto:bhfdsc@hdruk.ac.uk)

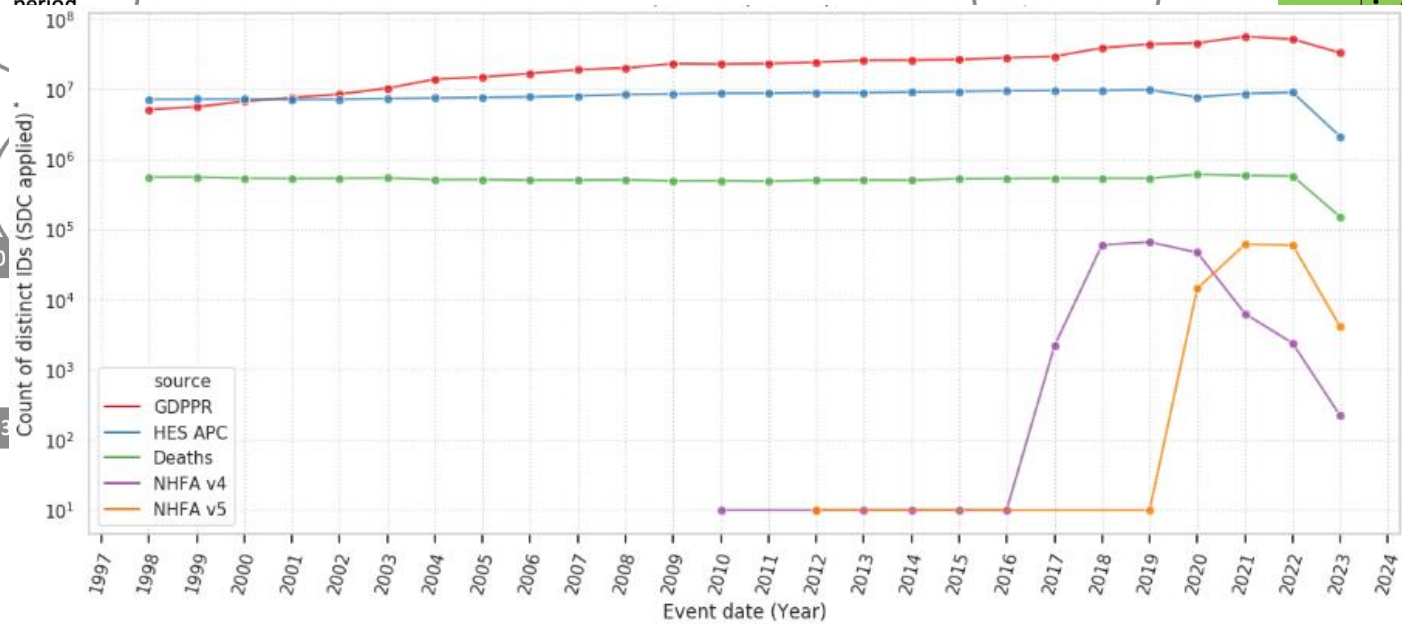


@HDRUK

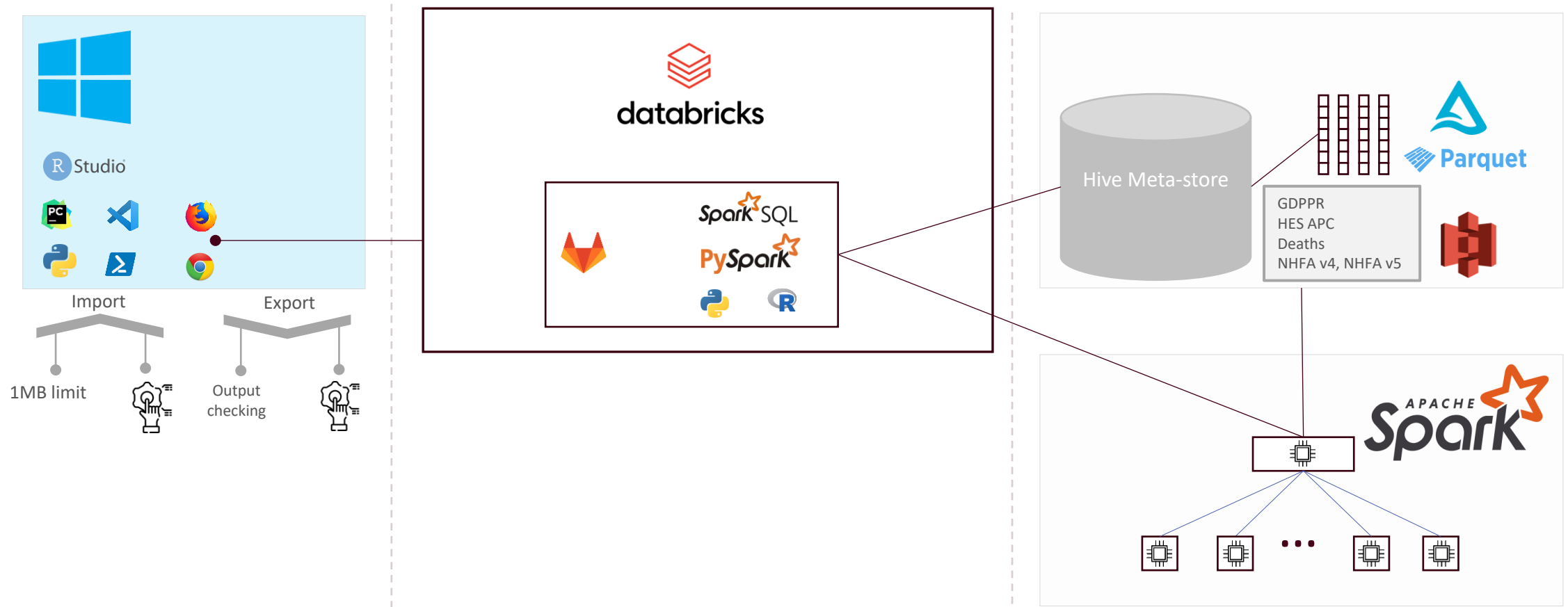
# DATA SOURCES ON COVID IMPACT



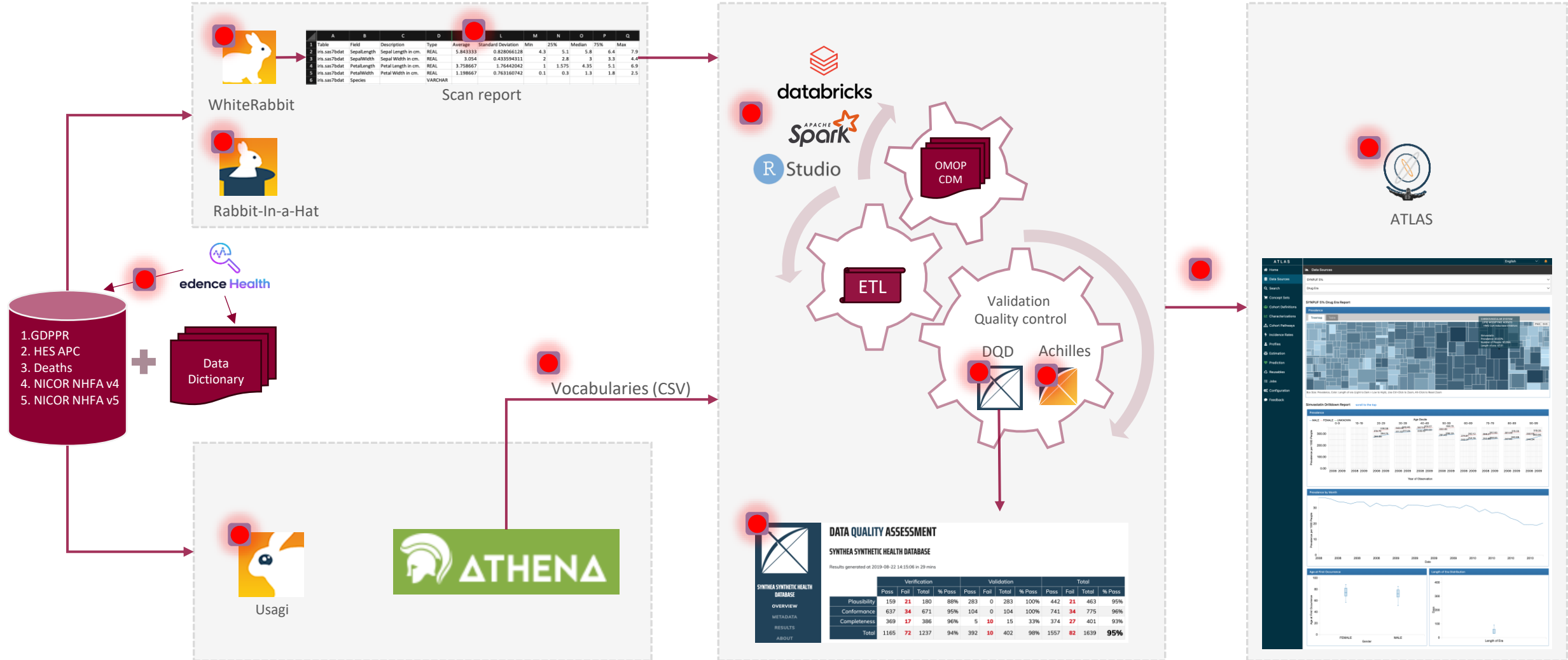
WALES / 3.2 million	
SAIL Databank	
16 users / 7 institutions	
Requested	35 requested / 30 provisioned
• ONS COVID-19 Infection Survey available, subject to approval	
• General Practice Monthly/Daily COVID	
Respiratory Infections - (2000-2020)	• Critical Care Dataset
Case - (2001-2021)	• Emergency Department Daily/Monthly
	• Outpatient Dataset for Wales
	• Outpatient Referral Dataset
	• Patient Episode Dataset
Immunisation (2001-2021)	• LIMS (Pillar 1, 2, 3)
	• ONS COVID-19 Infection Survey*
	• Test, Trace & Protect
	• Shielded People
	• Variant strain data (COG-UK)*
	• Covid Vaccination Dataset
	• Annual District Death Daily/Monthly
	• Consolidated Death Data Source
	• SNARC Quarterly/Weekly COVID
	• I/A
	• Wales Dispensing Dataset
	• NICOR Audits and Registers (pending approvals)
	• QIP Stroke Audit (pending approvals)
	• VR (pending approvals)
	• Annual District Birth Extract
	• Care Homes Index
	• Paternity Indicators Dataset
	• Congenital Anomaly Register (CARIS)
	• National Community Child Health
	• ONS Census (2011)* ONS Census (2021)*
	• Referral to Treatment Times
	• AIL Dementia e-Cohort
	• Welsh Ambulance Service Dataset
	• Wales Results Reporting Service
	• Welsh Demographic Service



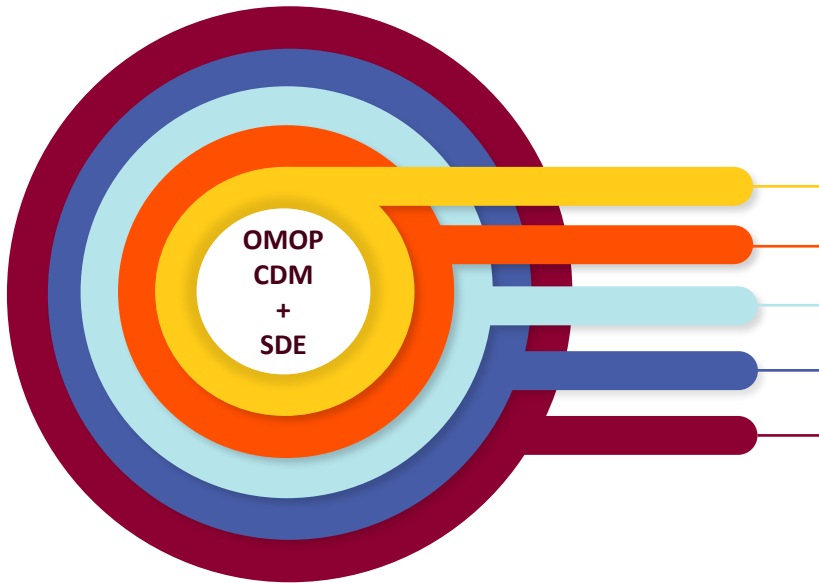
# NHS England Secure Data Environment



# ETL



# OMOP CDM in SDE



> Expand  
 > CRAN, PyPI, Conda  
 > Single database (schema)  
 > Dependencies  
 > Data structures  
 > Naming convention  
 > ETL Quality improvement  
 > Memory issues  
 > Federated studies  
 > Data input/output  
 > Studio Desktop vs Server  
 > UK Real-World Evidence Network  
 > Read-only  
 > Share learning  
 > CRAN, PyPI, Conda  
 > SPARK-ready OMOP CDM ETL  
 > OMOP CDM in Secure Data Environments  
 >> Git do primary key + Window

## Project members

Mehrdad Mizani, Jadene Lewis, Rouven Priedon, Thomas Bolton, John Nolan, Angela Wood

## Acknowledgements

We gratefully acknowledge the support of the team at edenceHealth NV (<https://edence.health>), including Silvia Jimenez, Shirah Cashriel, Emma Gesquite, and Anne Li, who provided invaluable collaboration in ETL development and guidance on OMOP CDM and EHDEN milestones. We extend our gratitude to Maxim Moinot and Daniel Prieto-Alhambra for their valuable insights which were instrumental in achieving the EHDEN reports.

This work was carried out with the support of the BHF Data Science Centre, led by Health Data Research UK (BHF Grant no. SP/19/3/34678).

This project is supported by the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement no 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

This work was supported by Health Data Research UK which receives its funding from HDR UK (HDRUK2024.0336) funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation (BHF) and Cancer Research UK.



[bhfdatasciencecentre.org](https://bhfdatasciencecentre.org)



[bhfdsc@hdrug.ac.uk](mailto:bhfdsc@hdrug.ac.uk)



[github.com/BHFDSC](https://github.com/BHFDSC)



[@BHFDataScience](https://twitter.com/BHFDataScience)

