



Better Care Insight Sharing Day

24 February 2022

Introduction and welcome



Objectives for today

1. Provide updates on ongoing work across the Better Care network:

- Policy and Insight Catalyst Workstream: What actions are required to help develop and scale learning health systems?
- Reflections on using health data for biomedical research
- CogStack: Fundamental Infrastructure for 'Unlocking' Electronic Health Record (EHR) data for clinicians, academics and population health analysis

2. Consider how the Progress and Impact Framework can support best practice research

3. Discuss wider resources and support available through HDR UK

- Patient and Public Involvement and Engagement for Better Care
- ALLEVIATE and DataMind Hubs

Agenda:

Time	Session	Who	Aims
Better Care Insight Sharing			
09.30	Introduction and welcome	Alastair Denniston , Director of INSIGHT, Associate Director HDR UK Midlands, Consultant Ophthalmologist, Birmingham	Introduction and overview of day
09.40 (30 min)	Overview of the Progress and Impact Framework (PIF) PIF for Better Care	Victoria Platt , Executive Director of Business Operations HDR UK Kate Sanders Wilde , Continuous Improvement and Business Systems Manager	<ul style="list-style-type: none"> What is the PIF? Why do we collect the information and how is it used? How can the PIF add value for all members of the Better Care community? How can we use the PIF over the next year to monitor our impact and delivery on our priorities?
10:10 (75 min)	What actions are required to help develop and scale learning health systems?	Tim Horton , Associate Director Tom Hardie , Improvement Fellow, The Health Foundation Nell Thornton-Lee , Improvement Analyst, The Health Foundation	<ul style="list-style-type: none"> Update from the Policy and Insight Catalyst workstream Share survey findings Discuss report recommendations
11:25	BREAK		
11.40 (45 min)	Hub update: ALLEVIATE DataMind	Emily Jefferson , Director of the HDR Alleviate Pain Data Hub, University of Dundee Jenni Harrison , Deputy Director of the HDR Alleviate Pain Data Hub, University of Dundee Ann John , Principal Investigator and Co-Director of DATAMIND, Swansea University Rob Stewart , Co-Director of DATAMIND, Kings College London	<ul style="list-style-type: none"> Update on new data resources and tools available to support research in pain and mental health through the hub network
12:25 (30 min)	Patient and Public Involvement and Engagement for Better Care	Sinduja Manohar , Public Engagement and Involvement Manager Rosanna Fennessey, National Better Care PPIE representative Michaela Regan, Better Care PPIE representative	<ul style="list-style-type: none"> Strategies and resources to support PPIE in Better Care

Agenda:

Time	Session	Who	Aims
12:55	LUNCH		
14:00 (60 min)	Reflections on using health data for biomedical research	<p>Chair: Anthony Brookes, University of Leicester</p> <p>Paul Lambert, Professor of Biostatistics, University of Leicester.</p> <p>Keith Abrams, University of Warwick</p> <p>Olalekan Lee Aiyegbusi, University of Birmingham</p> <p>Alastair Denniston, Director of INSIGHT, Associate Director HDR UK Midlands, Consultant Ophthalmologist, Birmingham</p> <p>Theo Arvanitis, University of Warwick.</p>	<ul style="list-style-type: none"> • Limitations and challenges of using health data for research use • Consider the key issues including the impact of standardised data models • Explore the scale of the problem • Illustrate potential defensive strategies that help to mitigate the risks
15:00 (15 min)	CogStack: Fundamental Infrastructure for 'Unlocking' Electronic Health Record (EHR) data for clinicians, academics and population health analysis	Tom Searle , King's College London	<ul style="list-style-type: none"> • Introduction to CogStack • Current impact and future plans for the platform • How to get involved and use the technology
15:15 (5 min)	Closing remarks and next steps	<p>Alice Turnbull, Programme Director, HDR UK</p> <p>Kevin Dunn, HDR UK Programme Manager, HDR UK Midlands</p>	<ul style="list-style-type: none"> • Wrap up, next steps and close
15:20	CLOSE		

Progress and impact framework


Victoria Platt, Executive Director Business Operations

Kate Sanders-Wilde, Continuous Improvement and Business Systems Manager

17/03/2022



Plan for the session

- What is the Progress and Impact Framework?
 - How does it fit into the wider HDR UK strategy
 - Why do we have one; how do we use it?
 - PIF: Better Care data
 - PIF for 2022/23
- 

Progress and Impact Framework

The progress and impact framework (PIF) provides a balanced view of the organisation's progress against the vision, goals and outcomes agreed with the core funders and HDR UK board.



Within 5 years we aim for the UK to be the most impactful place to do health data science

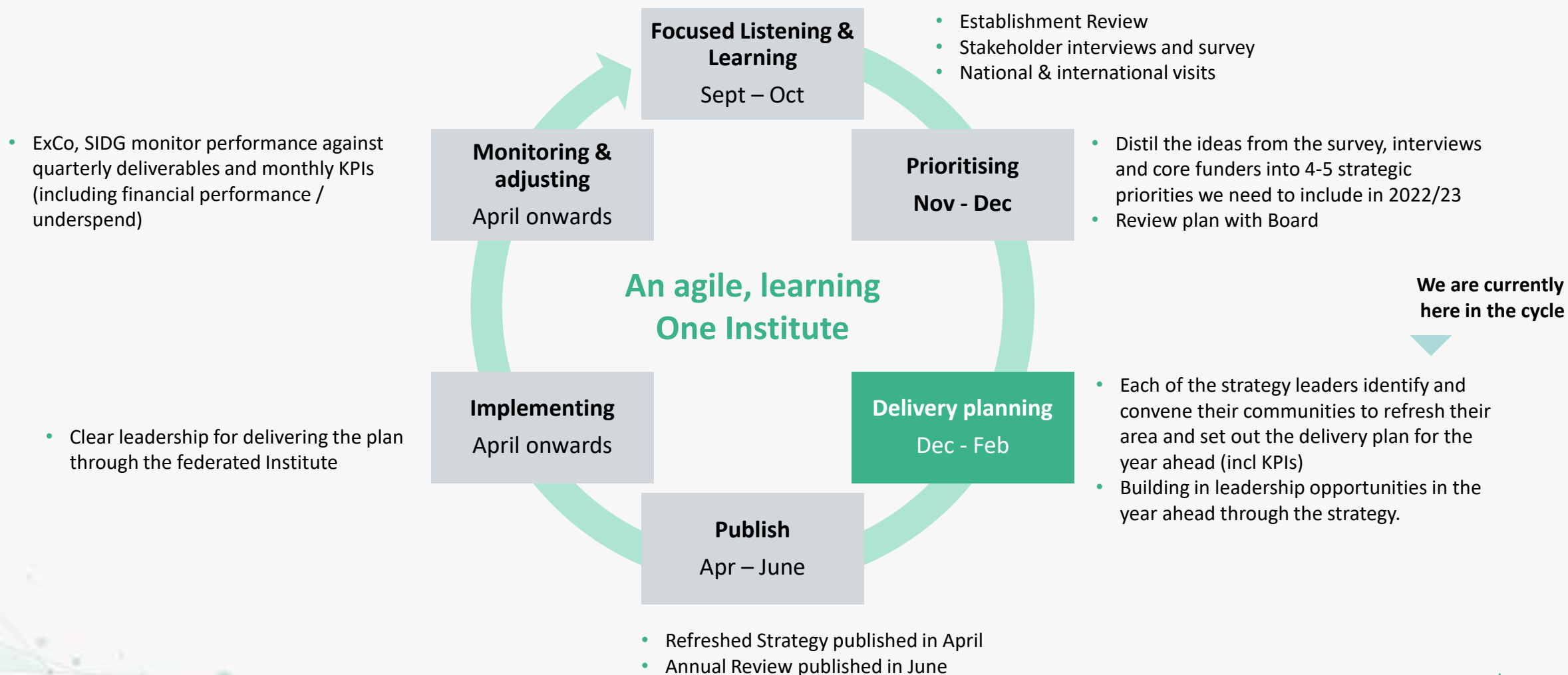
Vision		
20-year vision is for large scale data and advanced analytics to benefit every patient interaction, clinical trial, and biomedical discovery and to enhance public health		
Areas of focus	Goals (April 2018 to March 2023)	Outcomes (April 2018 to March 2023)
Uniting health data Alliance and Gateway	<ul style="list-style-type: none"> Delivering the Gateway; fundamental to the world's health data research, trusted by patients, public and practitioners Creating the Alliance with members from all the UK's major health data custodians* Establishing standards: Participation, Information Governance, Access, benefit sharing, and Trusted Research Environments Providing training for the infrastructure 	<ol style="list-style-type: none"> Efficient, safe access to large scale, diverse data for <i>researchers and innovators</i> Transparency of use of health data to <i>patients and the public</i>
Improving health data – Tools, methods and hubs	Creating better quality data in the Alliance and Gateway through: <ul style="list-style-type: none"> Tools and methods to measure and improve the data quality in the Gateway, including applied analytics and the human phenome 8-10 hubs improving the data 	<ol style="list-style-type: none"> Better data for <i>researchers and innovators</i>
Using health data Research discoveries and skills	<ul style="list-style-type: none"> Delivering UK-wide research programmes: Understanding Causes of Disease, Improving public health, Better Clinical Trials and Better Care Building health science user community: patients, public, academia, NHS, charities, and government Demonstrating major impact use cases Delivering training programmes and career pathways for health data scientists 	<ol style="list-style-type: none"> Better, more useful, research for <i>funders and public</i> – that no single research organisation could achieve alone
One Institute - Public trust & benefits, governance, team science and communities	<ul style="list-style-type: none"> Recognised internationally as one of the world's leading health data science institutes Scalable, trusted business model Inclusive, team-oriented culture built on the values of transparency, optimism, respect, courage and humility Successful QQR and leverage funding secured 	<ol style="list-style-type: none"> UK recognised as the most impactful place to do health data science

* Private company data strategy TBD

HDR UK's 12 delivery priorities for 2021/22

Strategy area	2021/22 delivery priorities
Uniting health data Alliance and Gateway	<ol style="list-style-type: none"> 1. Gateway services (convenience, speed of access, metadata, concierge) – use of these services, happy customers & a vibrant ecosystem of partners built on patient, public and user involvement 2. Federation across national Trusted Research Environments for specific uses (eg Data & Connectivity) 3.a. Breadth of relevant datasets (eg more high-use datasets such as the Zoe Symptom Tracker, PRUK) and coverage (whole UK population). Availability of Industry datasets through partnerships (eg hubs) 3.b. Diversity of datasets to ensure an appropriate representation of the population.
Improving health data Tools, methods and hubs	<ol style="list-style-type: none"> 4. Evidenced improvements to data (ie successful Hub Milestone 2 & 3) and more linked data assets on the Gateway with accessible descriptions (BHF Data Science Centre, National Core Studies, ICODA) that are being used to meet national priorities and policy. 5. Developing data engineering capability to get data onboard, link it and improve it 6. Embedding data utility, TRE and federation standards
Using health data Research discoveries and skills	<ol style="list-style-type: none"> 7. New high-impact outputs and innovations from national priorities, demonstrating our distinctive approach, including the impactful cross-national priority and cross-disease approach 8. Training 5,000 cross-sectoral health data researchers 9. 2-3 high-impact driver projects that harness the value of industry 10. International driver projects (part of ICODA) that guide principles and best practice for international uses of data
One Institute Public trust, benefits, governance, team science and communities	<ol style="list-style-type: none"> 11. Joined-up brand strategy to unite the different components (Alliance, ICODA, Gateway, hubs, training), promote partners and reach public and research communities 12. Quinquennial 2 vision, financial strategy and organisation development in place to achieve it

Where we are in the strategy development process



Progress and Impact Framework

The PIF provides a view of progress through three lenses:

- 1) **Maturity matrix** - can the organisation show **qualitative** progress towards each of its desired outcomes?
- 2) **Deliverables** - has the organisation delivered the **tangible** aspects of the agreed goals?
- 3) **Quantitative metrics (Key Performance Indicators)** - can the organisation demonstrate **quantified** progress against its goals and desired outcomes?

Outcomes self-assessment matrix – this tells us how we are doing against our desired outcomes

Last updated Q3 2021-22

Complete Largely complete Status Q4 2020/21 Status Q1 2021/22 Status Q2 2021/22 Status Q3 2021/22

	HDR UK QQR desired outcomes	Current level of maturity for each outcome The assessment templates showing maturity criteria for each outcome can be viewed here					Impact & output examples from Q2 2021/22
		Level 1 - Awareness	Level 2 - Initiated	Level 3 - Defined	Level 4 – Managed	Level 5 - Optimised	
Uniting	1. Efficient, safe access to large scale, diverse data for researchers and innovators			Data Access requests remain relatively limited via gateway outside of Rapid Call projects	1/3 criteria complete, work in progress on all other criteria		Data Use Register standards - driving improvement in Alliance members' registers
	2. Transparency of use for patients and the public				1/3 criteria complete, work in progress on all other criteria	Work begun on all criteria	Recruitment of more diverse representatives and voices, promoting the role of the PAB. Implementing a joint programme of work with ONS and ADR UK around public engagement.
Improving	3. Better data for researchers and innovators	Largely complete, utility improvement from areas outside of health not routinely collected					
Using	4a. Better, more useful, research - Understanding Causes of Disease		Largely complete: broadening of engagement required across the full range of researches and innovators	Largely complete. Early stage of delivery for UCD strategy	2/3 criteria complete, work in progress on final criteria	Work begun on 3/4 criteria	
	4b. Better, more useful, research – Clinical Trials			Work complete on 2/5 criteria and begun on one further criteria			
	4c. Better, more useful, research – Public Health				Largely complete, integrated approach partial success	Work begun on all criteria	
	4d. Better, more useful, research – Better Care			All criteria complete or largely complete and ongoing		Largely complete, ongoing development for national network of sites	North and South West sites now joining the Midlands substantive site in demonstrating wider regional engagement.
	4e. Better, more useful, research - Training	Largely complete: further understanding required of training needs internationally	Largely complete: case studies ongoing	Started work on all criteria for Level 3	Work ongoing for 2/3 criteria	Work begun on 4/5 criteria	Our bite-size learning modules are unique and will be seamlessly integrated into the Gateway. Cohort 1 of our PhD programme are already beginning to demonstrate their impact
One Institute	5. UK recognised as the most impactful place to do health data science	Largely complete: further understanding required of best practices		Ongoing work on demonstrating effectiveness of approach	4/5 criteria currently met, work begun on scalable model	Work complete (but continually evolving) on 4/6 and begun on a further 2/6	QQ2 Frontiers meetings – engaging wide community in development of HDR UK Black internship programme 2021 stakeholder interviews

Status of Delivery Priorities

Last updated Q3 2021-22

>90% of plan
 75%-90% of plan
 <75% of plan

Strategy area	2021/22 delivery priorities	Responsible	Completion
Uniting health data	1. Gateway services (convenience, speed of access, metadata, concierge) – use of these services, happy customers & a vibrant ecosystem of partners built on patient, public and user involvement	CG, DS	
	2. Federation across national Trusted Research Environments for specific uses (eg Data & Connectivity)	CG, DS, BG	
	3.a. Breadth of relevant datasets (eg more high-use datasets such as the Zoe Symptom Tracker, PRUK) and coverage (whole UK population). Availability of Industry datasets through partnerships (eg hubs)	DS, BG	
	3.b. Diversity of datasets to ensure an appropriate representation of the population	DS, BG	
Improving health data	4. Evidenced improvements to data (ie successful Hub Milestone 2 & 3) and more linked data assets on the Gateway with accessible descriptions (BHF Data Science Centre, National Core Studies, ICODA) that are being used to meet national priorities and policy.	CG, DS, JL, AW, CSud	
	5. Developing data engineering capability to get data onboard, link it and improve it	BG	
	6. Embedding data utility, TRE and federation standards	BG	
Using health data	7. New high-impact outputs and innovations from national priorities, demonstrating our distinctive approach, including the impactful cross-national priority and cross-disease approach	RW	
	8. Training 5,000 cross-sectoral health data researchers	RW, SCad	
	9. 2-3 high-impact driver projects that harness the value of industry	RW	
	10. International driver projects (part of ICODA) that guide principles and best practice for international uses of data	AM	
One Institute	11. Joined-up brand strategy to unite the different components (Alliance, ICODA, Gateway, hubs, training), promote partners and reach public and research communities	AW	
	12. Quinquennial 2 vision, financial strategy and organisation development in place to achieve it	AB/AH, VP, CSmi	

Key Performance Indicators

Uniting and improving data – January 2022 + Q3 2021/22

¹  >90% of plan  75%-90% of plan  <75% of plan




² Change vs previous month for monthly metrics, previous quarter for quarterly metrics; improvement in green font, decline in red font
³ Total year to date is for full year 2021/22 for quarterly metrics; for monthly metrics; January 2022

Theme	Category	Performance Indicator	Metric	Reporting period	Actual ¹	Plan	Change vs last period ²	Total YTD ^{1,3}	YTD plan
Uniting and improving data	Datasets	Quantity of datasets (m)	Gateway datasets (cumulative)	January	724	1,125	-3	724	1,125
		Number of datasets publishers	Gateway dataset publishers (cumulative)	January	48	74	0	48	74
		Quality of metadata (m)	% Gateway datasets with technical metadata	January	59	58	0	59	58
		Dataset coverage	% simple Gateway searches with at least one result	January	99	90	0	99	90
		Features and Story points	Number of Features & Story Points delivered each period	Q3 21/22	180	594	-120	1,440	1,782
	Access and use	Gateway usage	Registered Gateway users (cumulative)	January	1,900	N/A	82	1,900	N/A
		Gateway usage	Access requests via Gateway	January	23	88	15	168	594
		Hub activity	Hub contracts	Q2 21/22	37	78	12	102	206
		Hub industry activity	Hub contracts with industry involvement	Q2 21/22	15	9	3	55	26
		Utility Framework evaluation	Datasets on Gateway with utility framework evaluation, %	Q2 21/22	18	25	0	18	25
		Cohort Discovery	Number of datasets searchable through cohort discovery	Q2 21/22	7	6	1	19	18
		Federation	Count of metadata catalogues through Gateway	Q2 21/22	1	2	1	1	6
	International	Reach	International datasets/tools	Q2 21/22	22	N/A	8	22	N/A
			International Alliance partners	Q2 21/22	19	N/A	0	19	N/A
	Financial	Central spend	Spend on Gateway and Programmes, £'000	January	4,152	4,152*	-8,113	12,149	12,149*
		Distributed spend	Spend by National Priorities and Hubs, £'000	Q2 21/22	365	365	-1,711	4,631	4,631

*note that figures are based on financial reforecasting hence 'Actual' is equivalent to 'Plan'

Key Performance Indicators

Using data and One Institute data – January 2022 + Q3 2021/22

¹  >90% of plan  75%-90% of plan  <75% of plan

² Change vs previous month for monthly metrics, previous quarter for quarterly metrics; improvement in green font, decline in red font

³ Total year to date is for full year 2021/22 for quarterly metrics; for monthly metrics; January 2022

Theme	Category	Performance Indicator	Metric	Reporting period	Actual ¹	Plan	Change vs last period ²	Total YTD ^{1,3}	YTD plan
Using data	Impact	Publications	Peer-reviewed publications, authored by research community	January	45	33	7	304	334
			Publications, authored by central team	January	1	1	1	6	10
		Repositories	New open source digital tools openly accessible on platforms	January	5	7	-4	52	76
	Training	Delivery of professional development programme	Total number of trainees	Q3 21/22	1,311	2,813	-2,288	11,102	7,125
	Financial	Central spend	Spend on centrally delivered programmes, £'000	January	28	28*	1	947	947*
		Distributed spend	Spend by National Priorities, Training programmes, £'000	Q3 21/22	1,731	1,731*	-160	5,291	5,291*
One Institute	Reach	HDR UK website usage	HDR UK website users	January	25,648	34,129	-2,650	263,254	276,714
		HDR UK media prominence	HDR UK media mentions	January	14	200	-74	1,910	2,000
		HDR UK audiences	HDR UK reach	January	7,222	7,352	0	7,222	7,352
		HDR UK events	Central events participants	Q3 21/22	1,747	4,513	567	6,251	12,825
	Community	HDR UK community membership	HDR UK community members (cumulative)	Q3 21/22	1,147	1,481	8	3,070	3,260
	Trust	Patient/Public involvement	PPIE participants	Q3 21/22	4,592	1,000	3,045	8,999	3,000
	People	Central team retention	Number of staff members who have left HDR UK employment	Q3 21/22	7	N/A	1	16	N/A
		Recruitment	Recruitment conversion rate, %	Q3 21/22	90	N/A	-10	90	N/A
	Financial	Central Spend	One Institute programmes and Central staffing, £'000	January	308	308*	-20	2,696	2,696*
		Funding	Funds under management, £m	Q3 21/22	147	N/A	0	147	N/A

*note that figures are based on financial reforecasting hence 'Actual' is equivalent to 'Plan'

Performance and Impact Framework - Output

HDR UK uses the PIF to help it continually improve and increase its likelihood of achieving its goals, outcomes and vision. It is used monthly, quarterly, annually and as part of the Quinquennial Review process.

Performance is reported using the PIF to the following meetings:

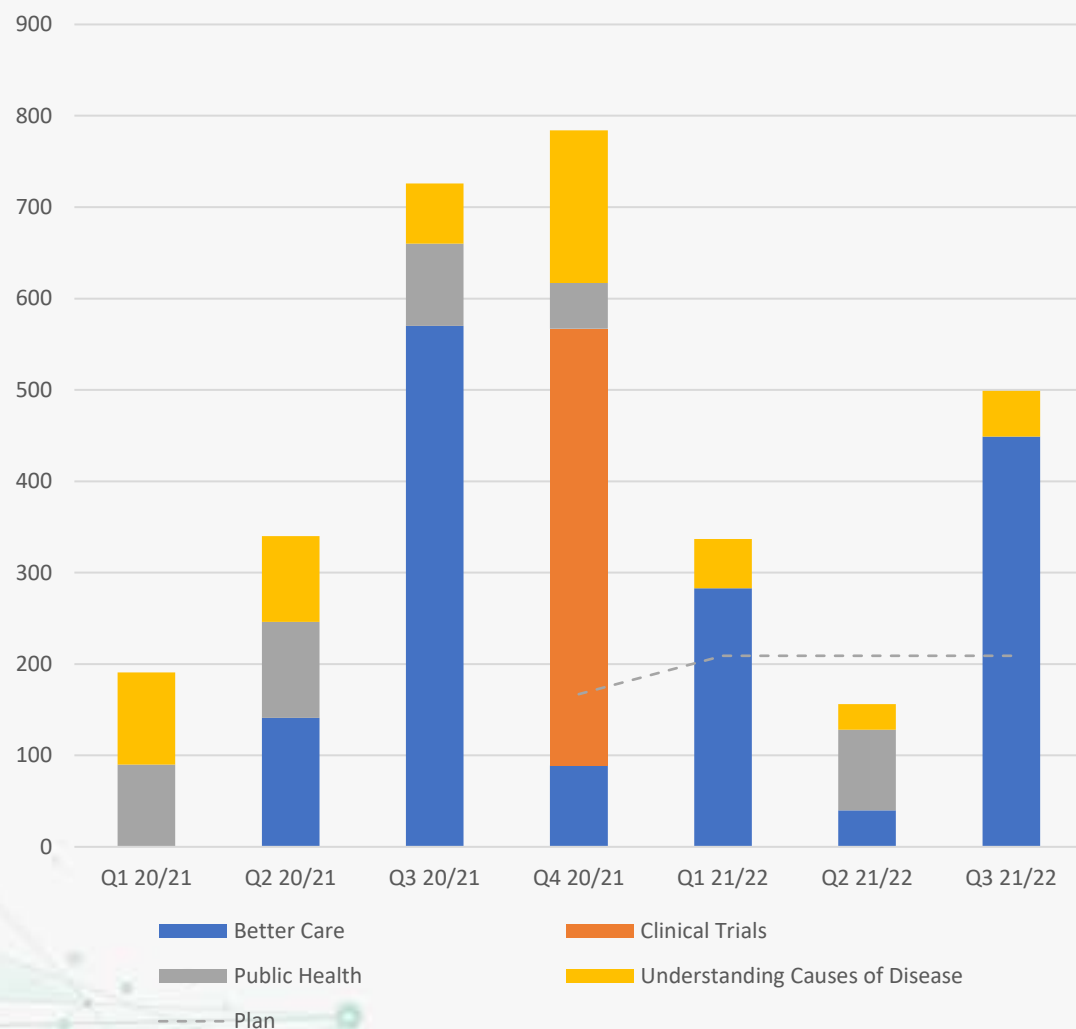
- **Monthly:** KPIs to ExCo
- **Quarterly:** KPIs + Deliverables + Outcome Assessment to ExCo, SIDG and Board
- **Annually:** KPIs + Deliverables + Outcome Assessment to ExCo, SIDG, Board and Core Funders

Better Care data

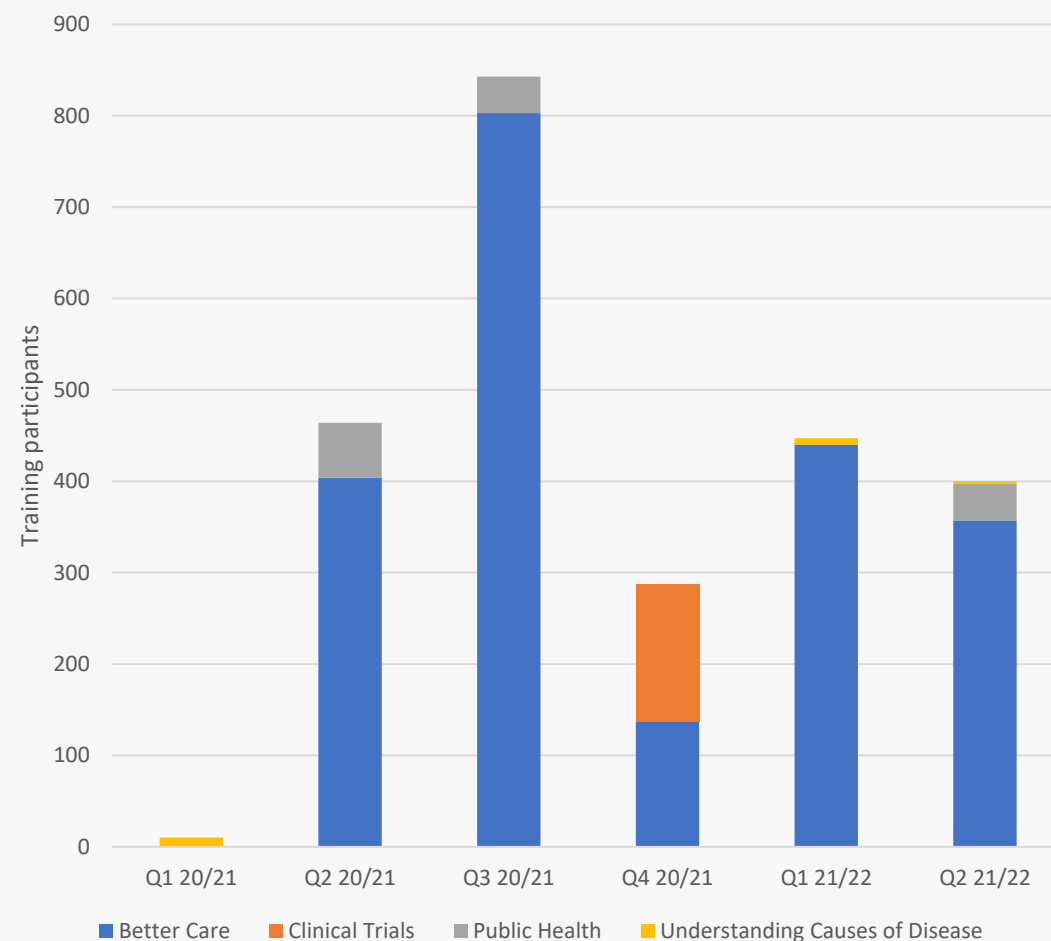
April 2020 – September 2021



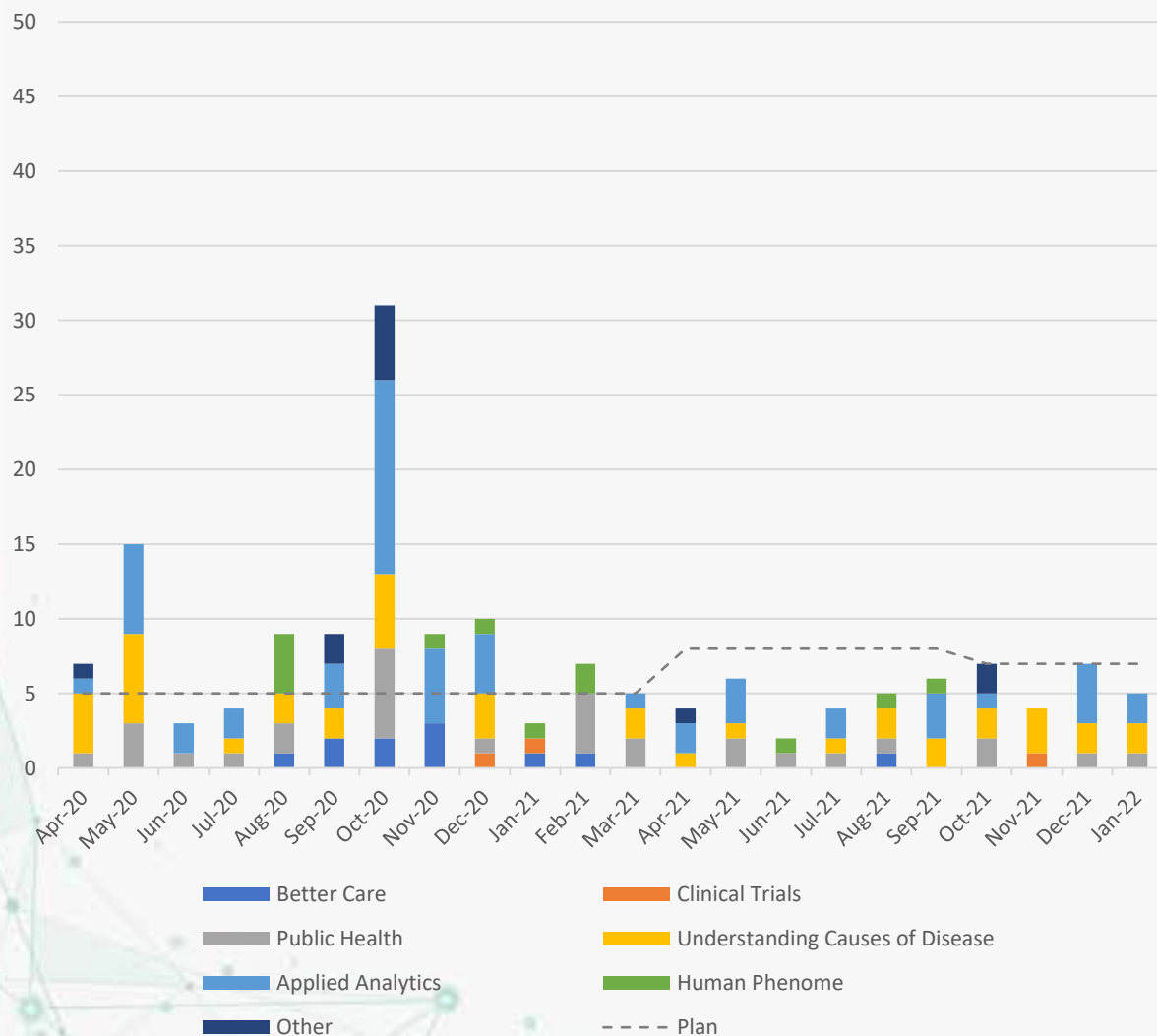
Using PPIE Participants



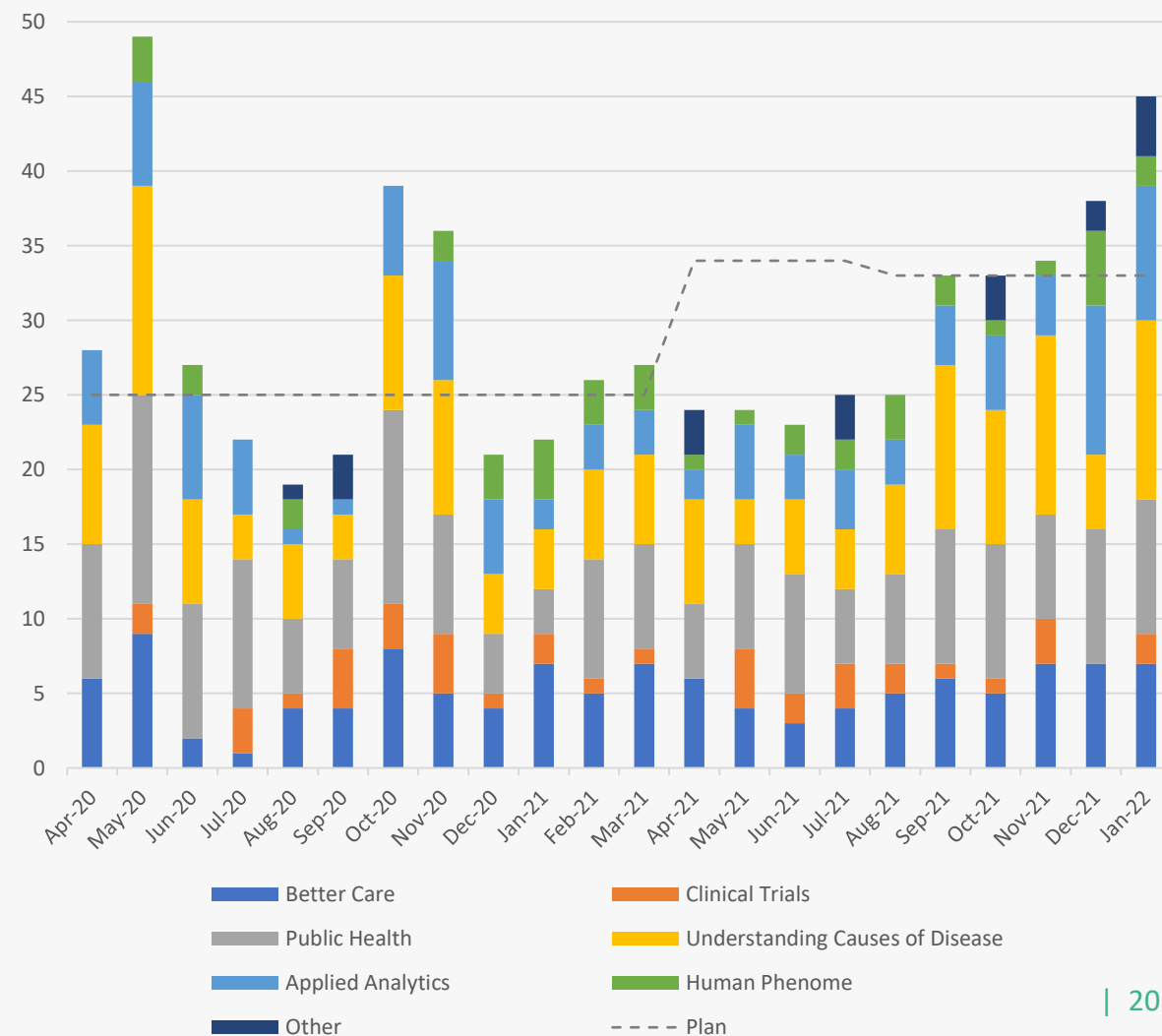
Using Training Participants



Github repositories added



Open source publications



Questions

Discussion: How can we make this as useful as possible?

- Delivery planning – how best to represent the work of the National Priorities?
- KPIs – which ones are most helpful?

What actions are required to help develop and scale learning health systems?

Tom Hardie, Improvement Fellow

@tlhardie1

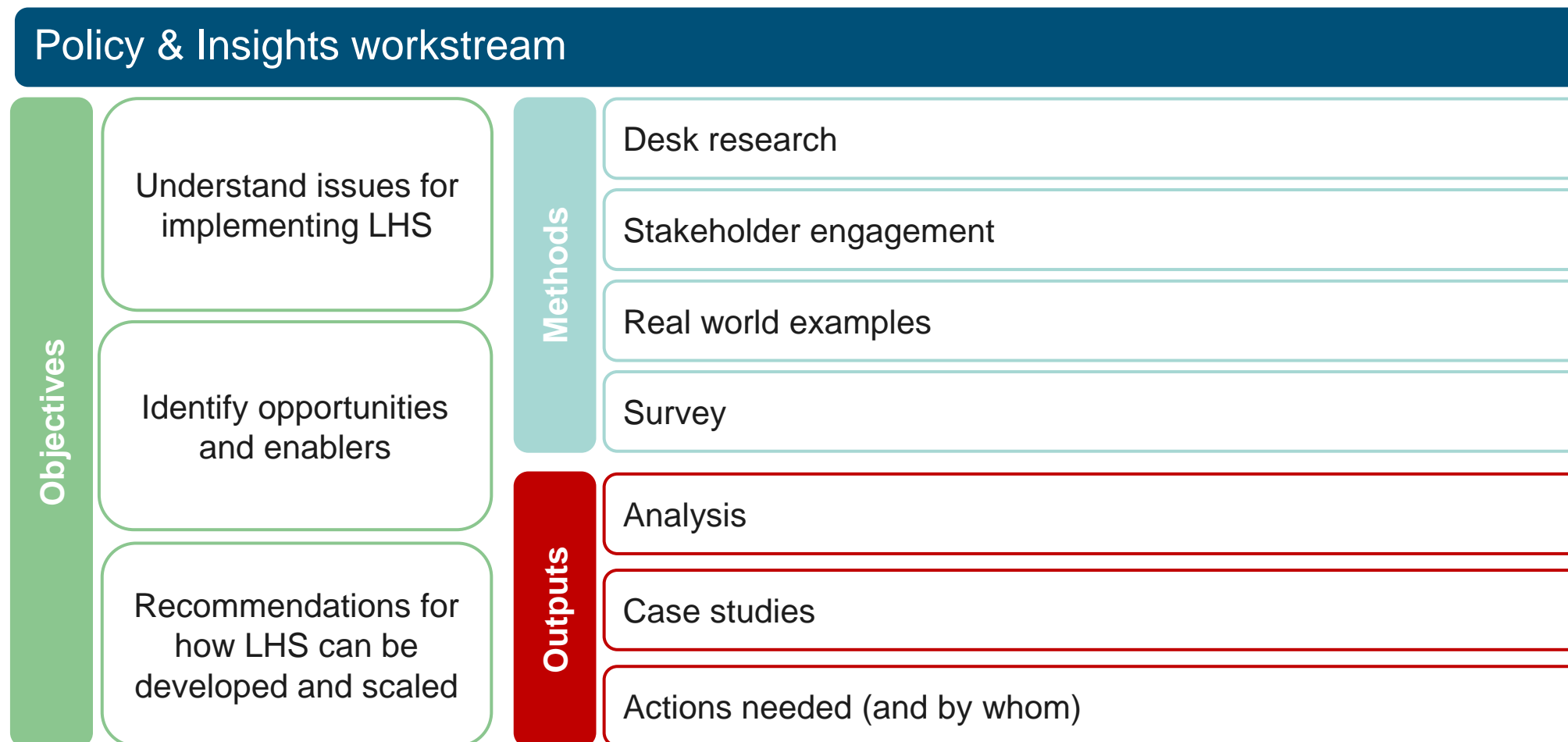
Nell Thornton-Lee, Improvement Analyst

@nellthorntonlee

24 February 2022



Workstream overview



What do we mean by learning health systems?

A set of **activities and assets** that can enable continuous learning and improvement of health and care services.

Learning health systems bring about change through **iterative learning cycles** based on:

1. generating and analysing **data**,
2. drawing knowledge from that data, and
3. formulating and testing **service changes**.

The scale and focus can vary



Focus: **Condition-based**



Focus: **Place-based**



Focus: **Thematic**



Nightingale Hospital
London

Focus: **Condition-based**

Scale

National

Regional

Local system

Organisational

Team/service

But there are common elements

While LHS vary in size, scope and complexity, they all have key activities in common:

- The **provision of services**
 - At the core of an LHS sits a service provider or providers – the aim is to improve service provision and outcomes
- A **learning community**
 - Multidisciplinary group who participate in the LHS, motivated by a common ambition of improving services and outcomes
- The **learning and improvement cycle**
 - Generating, analysing and drawing knowledge from data, and identifying and implementing improvements

LHS also require assets



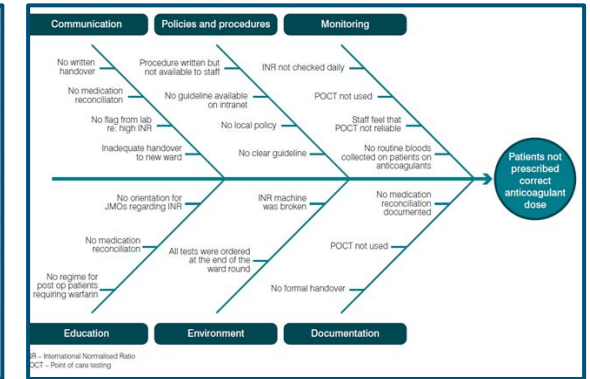
Data sources, platforms, systems and tools



Data analytics and research capability

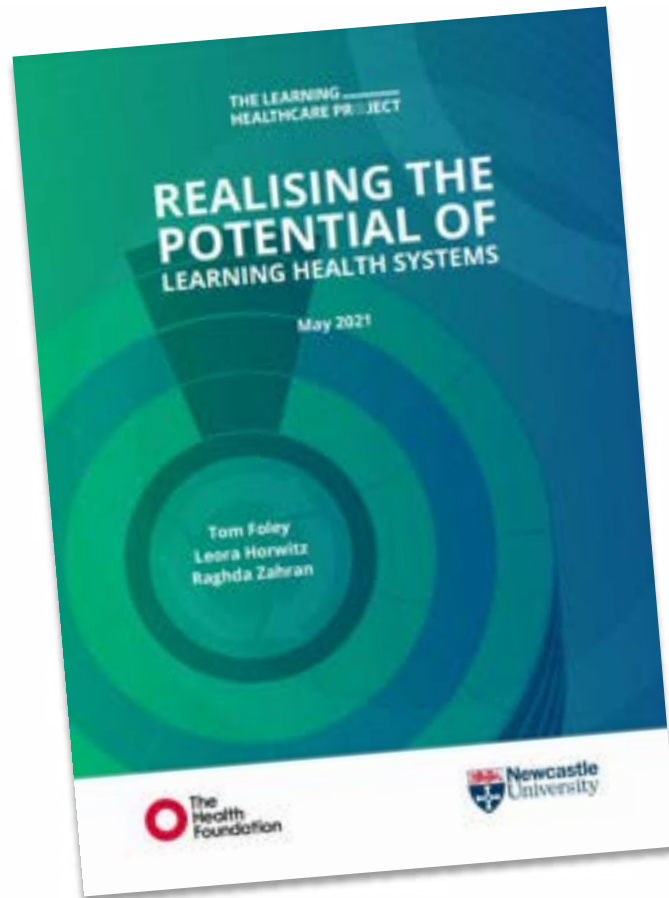


Learning communities and networks



Improvement capability and mechanisms

The work of a LHS is never done

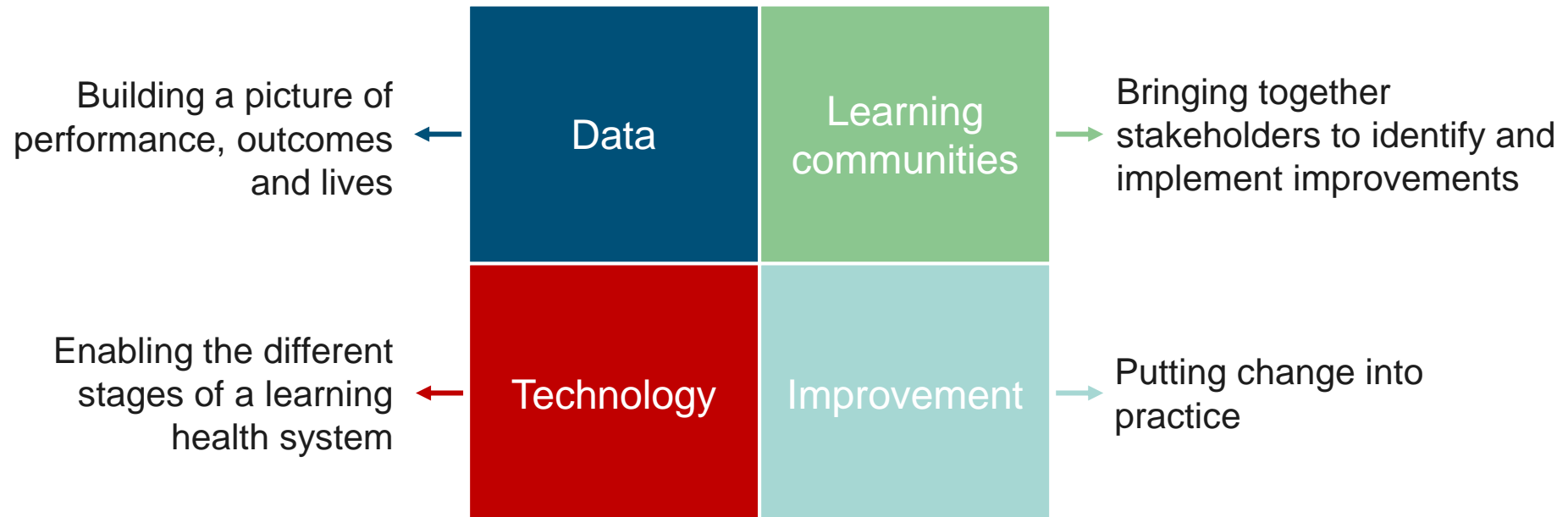


“ A learning health system is an **ongoing journey rather than a destination**; the very concept of a Learning Health System is that there is always something new to learn ”

Many potential benefits to outcomes & processes

Clinical effectiveness and safety			Improved use of technology
Reducing unwarranted variation			Staff engagement
Equity of access and outcomes			Patient-centred care
Efficiency and cost effectiveness			Personalisation
Research into practice			Diversity of perspectives in improvement process

Key lines of enquiry



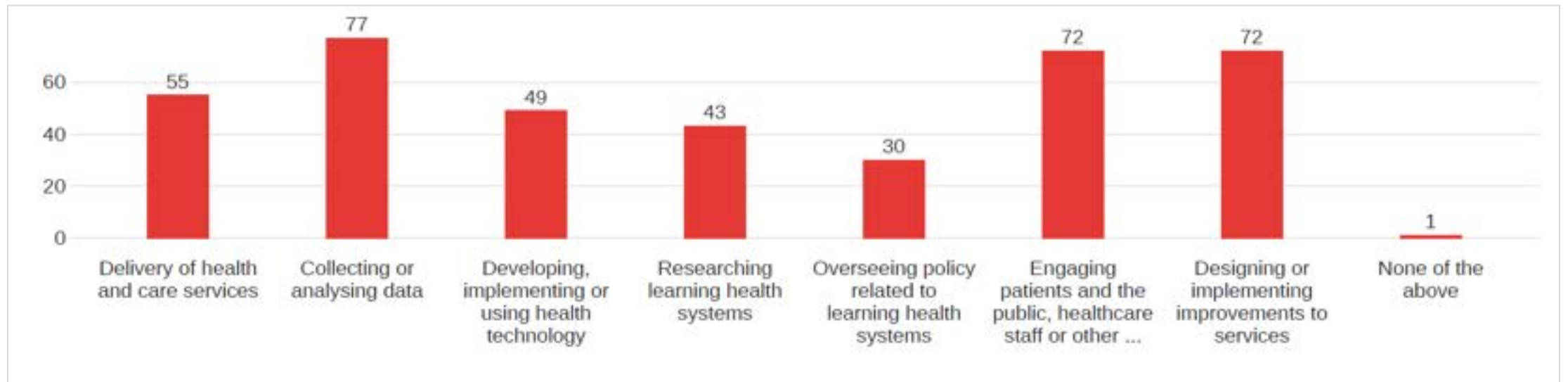
Survey objectives

- Through this survey we sought views on:
 - the **opportunities and challenges** for developing and scaling LHS the UK
 - the **key actions** that policy makers and health care organisations can take
- Developed options informed by literature review, stakeholder conversations, and case studies.
- All are important, but we wanted to identify **most important issues and actions**.

Who we asked

- Received 126 responses
- Senior stakeholders with expertise in one or more areas of Learning Health Systems from across UK

The following is a list of activities relevant to learning health systems. Which, if any, are you or have you recently been involved in?



Next steps

- Currently analysing our survey results
- Will be sharing our findings in a final report later this year
- Report will set out what actions are required to support the development and scale of learning health systems

Thank you

Tom.Hardie@health.org.uk

Nell.Thornton-Lee@health.org.uk

Tim.Horton@health.org.uk

Ellen.Coughlan@health.org.uk

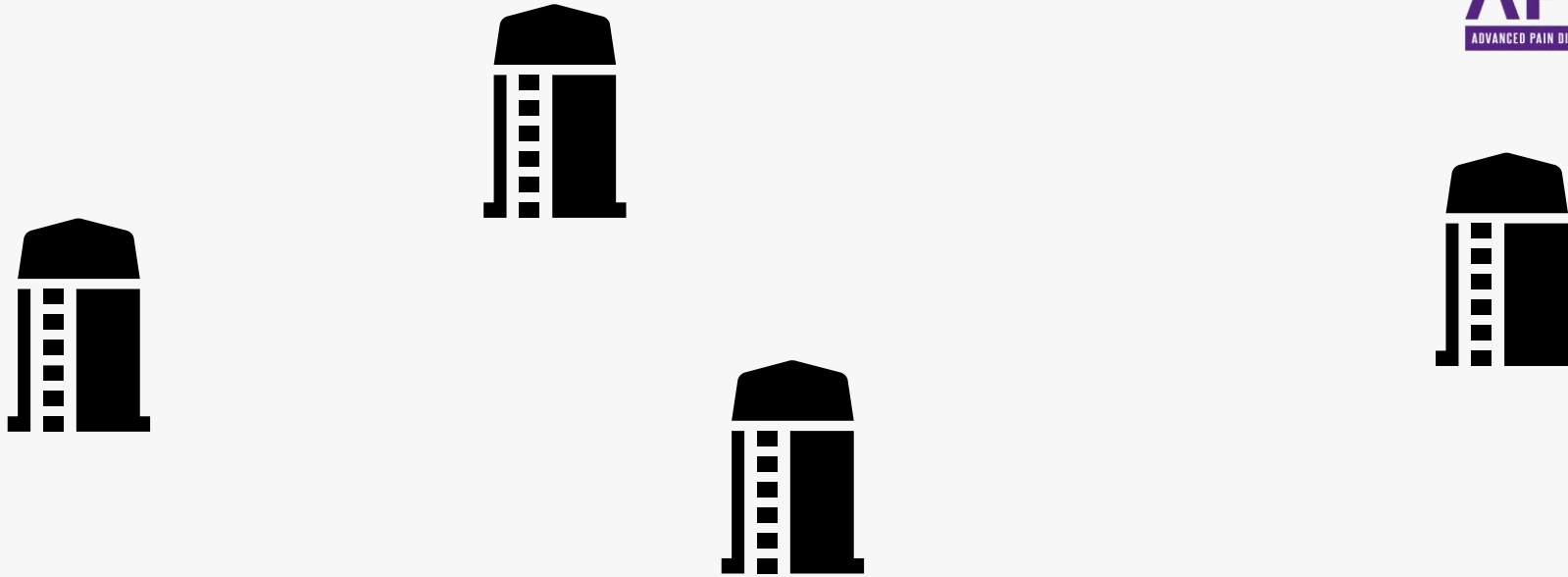
Alleviate: The Advanced Pain Discovery Platform (APDP) Data Hub

APDP consortia and data hub team 'kick off' meeting

17/03/2022 | Professor Emily Jefferson



Problem statement



- Many silos of datasets and pain disciplines which do not talk to one another
- Datasets are in different formats
- Datasets are known about via word of mouth

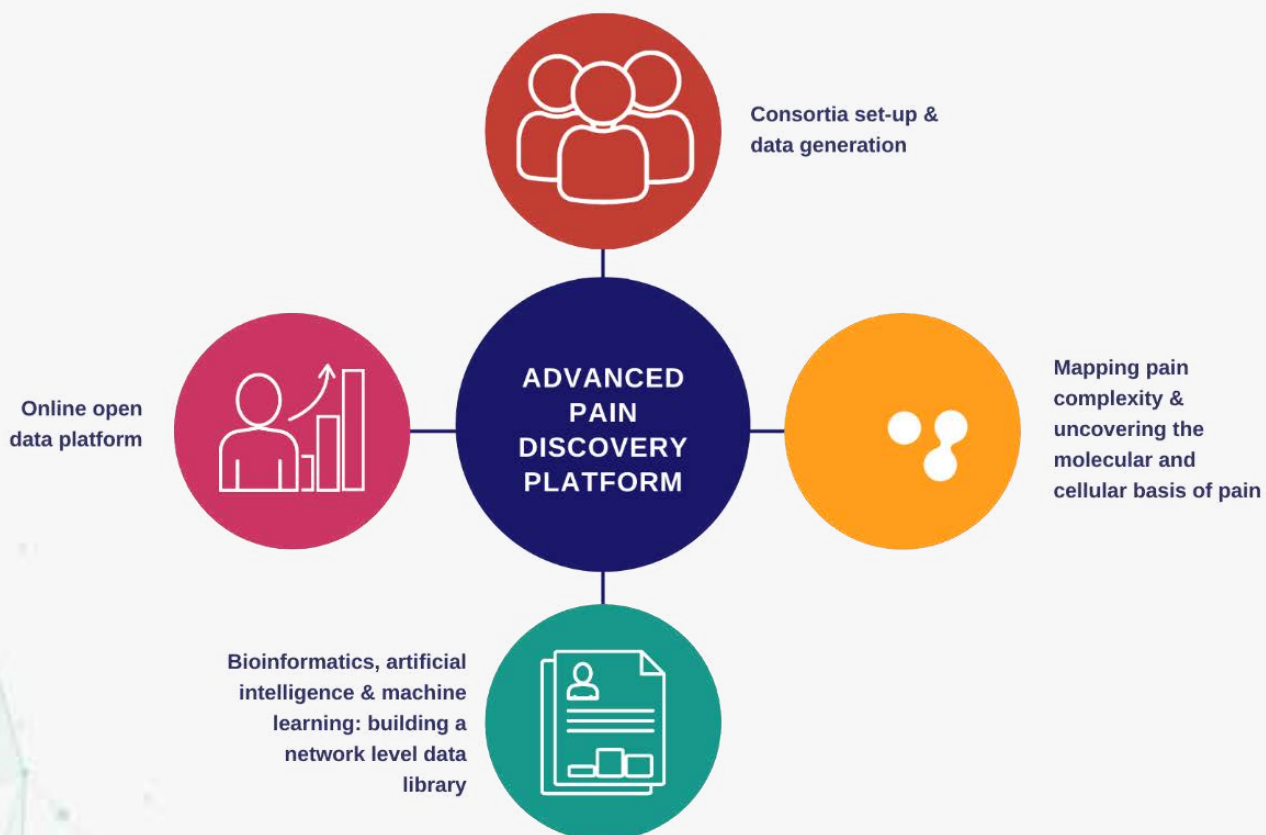


The Advanced Pain Discovery Platform
(APDP) Data Hub



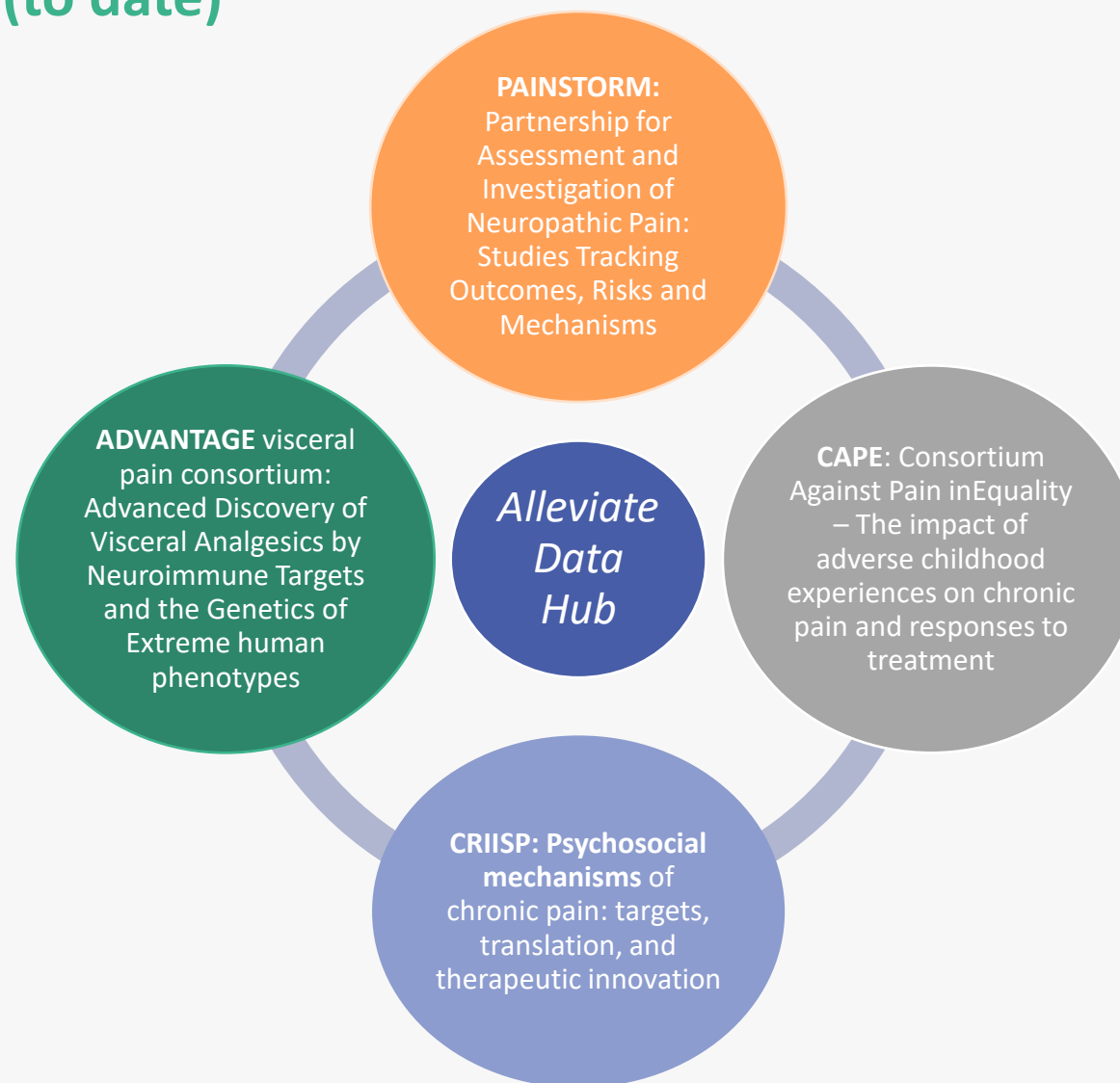
Vision: To transform UK pain datasets to be Findable, Accessible, Interoperable and Reusable (FAIR) link with expert data engineering, integrated into the Health Data Research (HDR) UK Innovation Gateway to enhance responsible, timely and trustworthy analysis by national and international researchers and innovators

The Advanced Pain Discovery Platform (APDP)



- 5 years - established initially with >£34M
- Funded through the Government's Strategic Priorities Fund and delivered in partnership through MRC, ESRC, BBSRC, Versus Arthritis and Eli Lilly
- Delivers a consortium-based platform of national scale that will break through the complexity of pain and reveal new treatment approaches to address a wide spectrum of chronic and debilitating clinical conditions.

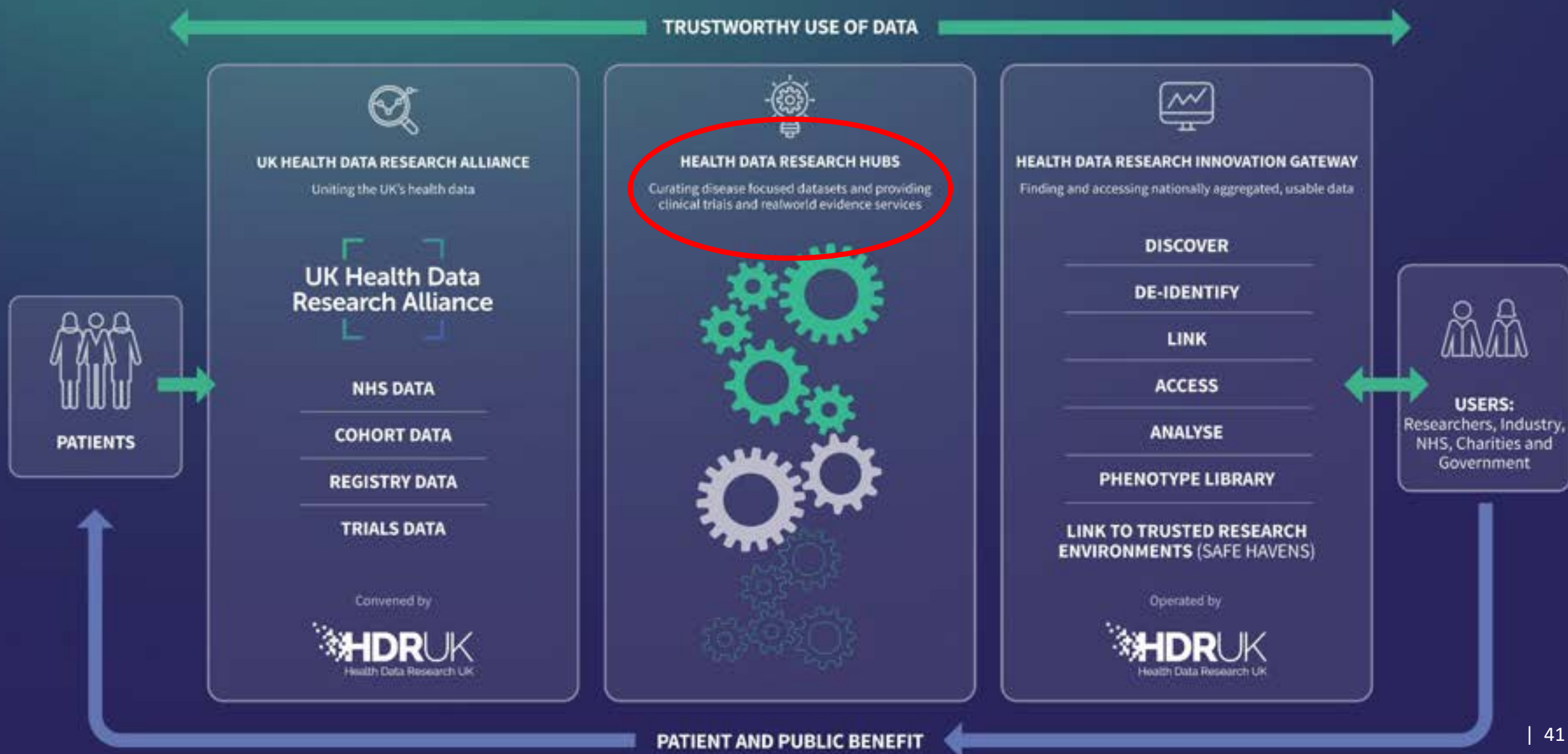
APDP Consortia (to date)



Alleviate is the data hub of the APDP, and will collaborate with the consortia to streamline access to data and on-board data into the Alleviate federated platform where appropriate

HDR UK's mission is to unite the UK's health data to enable discoveries that improve people's lives

UNITING THE UK'S HEALTH DATA TO MAKE DISCOVERIES THAT IMPROVE PEOPLE'S LIVES



Health Data Research Hubs

Centres of excellence with expertise and tools developing data to provide insights



**British Heart Foundation
Data Science Centre**

Led by Health Data Research UK



DATA MIND
The Health Data Research Hub for Mental Health

Discover-NOW

Health Data Research Hub for Real World Evidence



DIGITRIALS

Health Data Research Hub for Clinical Trials



Gut Reaction

Health Data Research Hub



PIONEER

Health Data Research Hub



BREATHE

Health Data Research Hub

DATA-CAN

The Health Data Research Hub for Cancer



INSIGHT

Health Data Research Hub



Alleviate

The Advanced Pain Discovery Platform
(APDP) Data Hub

Alleviate
The Advanced Pain Discovery Platform
(APDP) Data Hub

HDRUK
Health Data Research UK

APDP
ADVANCED PAIN DISCOVERY PLATFORM

**Health
Informatics
Centre**

HDR UK Gateway



←

→

↺

healthdatagateway.org

Health Data Research Innovation Gateway

Explore

Help

Usage Data

About Us

News

Community

Q

+ Add new

Sign in

Gateway to health data and tools for research

Search, discover and request access to hundreds of datasets, tools and resources for your research. Join the thousands of researchers and scientists worldwide who are already using the Gateway for research and scientific discovery.

Q

I'm looking for...

datasets 652	tools 151	projects 268	courses 195	papers 1,127	dataset requests 294	registered users 1,368	searches last month 9,127
-----------------	--------------	-----------------	----------------	-----------------	-------------------------	---------------------------	------------------------------

View all usage data.



← → ↻ phenotypes.healthdatagateway.org

HDRUK Phenotype Library

Home Phenotypes Concepts API About + Login

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

791 Phenotypes

1618 Concepts

106627 Clinical Codes

25 Data Sources

16 Coding Systems

A Reference Catalogue of Human Diseases

Connected. The Phenotype Library is accessible via an API to support interoperability, is integrated with health dataset information in HDR-UK's Innovation Gateway, and hosts content from numerous contributing organisations.

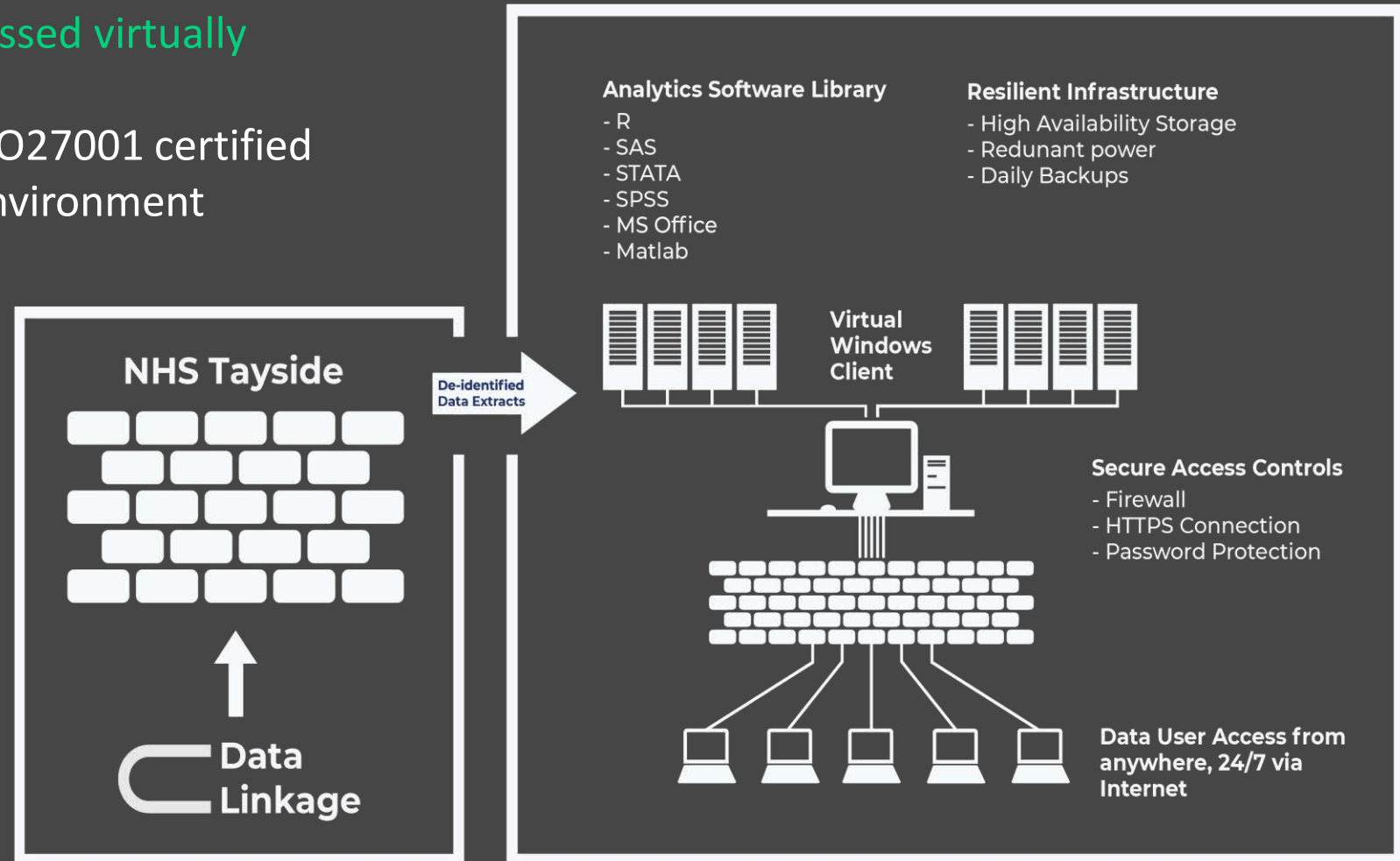
Patient-focused. The Library is enabling important research to improve patient health and well-being. Content spans major disease areas, including heart disease, cancer, COVID-19 and other common and rare human health conditions. Curated collections from contributors such as the HDR UK BREATHE Hub for respiratory health share clinical expertise to tackle critical research questions.

Cutting-edge. Built with a focus on computability, this resource aims to drive the next generation of research methods. Integration with *PhenoFlow* enables executable implementations of the phenotypes in our collection, while the API and R package client facilitate integration of the Library content directly into other analysis workflows.

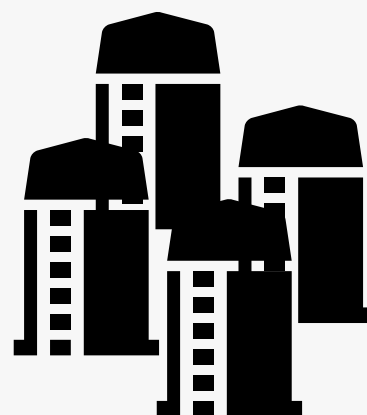
Accredited Safe Haven Environment/Trusted Research Environment (TRE)

Data can't leave and is accessed virtually

ISO27001 certified
environment



APDP
ADVANCED PAIN DISCOVERY PLATFORM



HDRUK
Health Data Research UK

Alleviate
The Advanced Pain Discovery Platform
(APDP) Data Hub

Alleviate Collaboration



Professor Emily Jefferson (PI):
Health Data Science



Professor Lesley Colvin:
Pain Medicine



Antony Chuter:
PPIE Co-lead



Dr Philip Quinlan:
Head of Digital Research
Service



Professor Frances Williams:
Genomic Epidemiology



Professor Dorothee Auer:
Neuroimaging



Professor Ana Valdes:
Musculoskeletal Genetics



Professor Andrew Rice:
Pain Research



Professor Blair Smith:
Pain Medicine



Richard Walls:
Business Development &
Relationship Manager



Professor Timothy Hales:
Anaesthesia



Professor Edmund Keogh:
Psychology



Professor Irene Tracey:
Anaesthetic Neuroscience



Professor David Bennett:
Neurology & Neurobiology



Professor Weiya Zhang:
Epidemiology



Christopher Hall:
Alleviate Chief Data Officer



Professor Victoria Chapman:
Neuropharmacology



Jillian Beggs:
PPIE Co-lead

Alleviate New Team



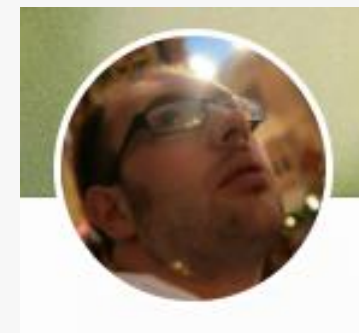
Dr Jenni Harrison
Alleviate Deputy Director



Scott Horban
Data Engineer



Richard Walls:
Business Development &
Relationship Manager



Christopher Hall:
Alleviate Chief Data
Officer

Alleviate
The Advanced Pain Discovery Platform
(APDP) Data Hub

HDRUK
Health Data Research UK

APDP
ADVANCED PAIN DISCOVERY PLATFORM

**Health
Informatics
Centre**

Hybrid Model of Data Discovery, Management and Provisioning

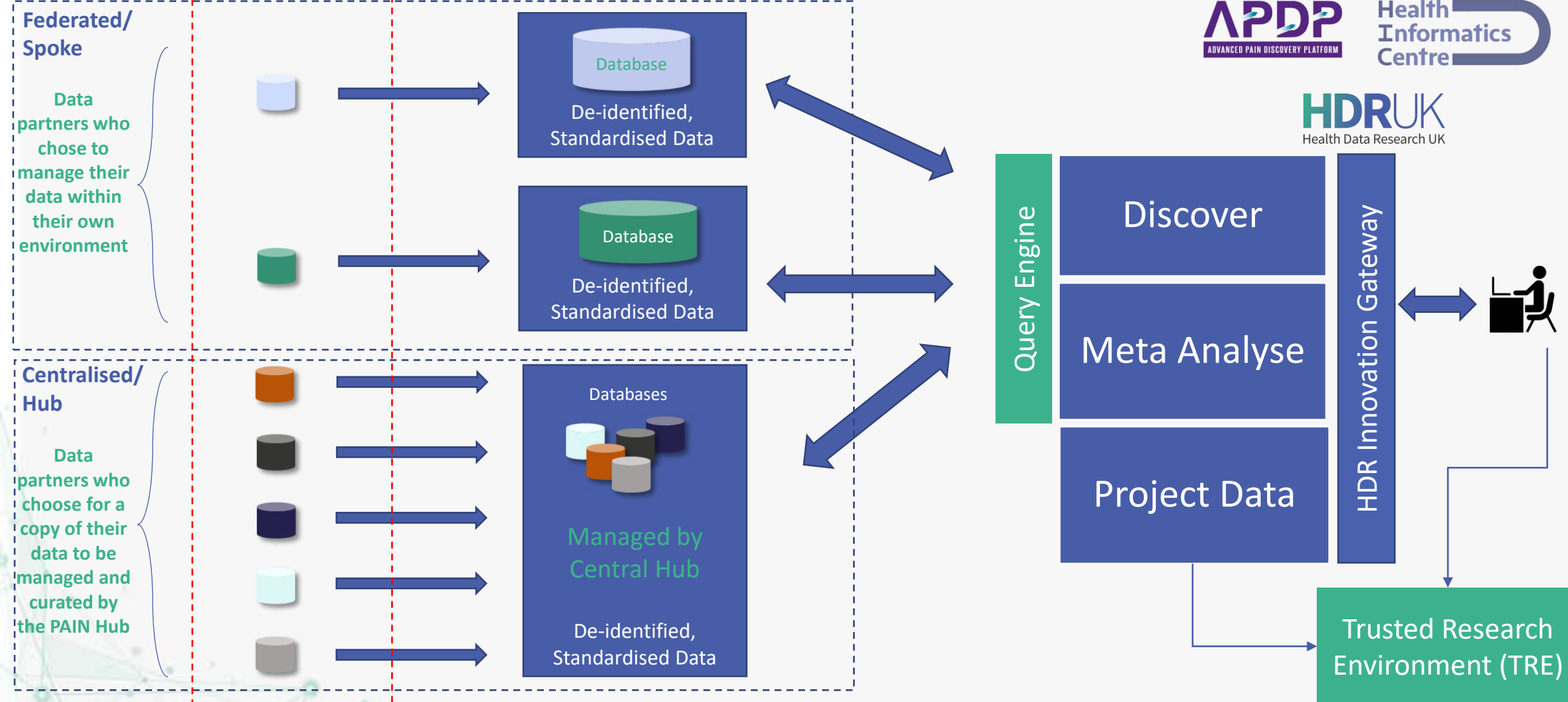
Alleviate
The Advanced Pain Discovery Platform
(APDP) Data Hub


HDRUK
Health Data Research UK

APDP
ADVANCED PAIN DISCOVERY PLATFORM


Health Informatics Centre

HDRUK
Health Data Research UK



CO-CONNECT

HomeMeet the TeamData Standards




COVID - Curated and open analysis and research platform

What is CO-CONNECT?

CO-CONNECT is a multi-million pound research project to help scientists across the UK to access the data they need more easily to help develop potential therapies and treatment for COVID-19. It is funded by the [Medical Research Council](#) (Part of UKRI) and the [Department of Health and Social Care](#) (part of NIHR) as part of their direct response to the Pandemic.

More information on our UKRI Funding



Cohort Discovery on the HDR Innovation Gateway

HDRUK
Health Data Research UK

healthdatagateway.org

Hybrid Model of Data Discovery, Management and Provisioning

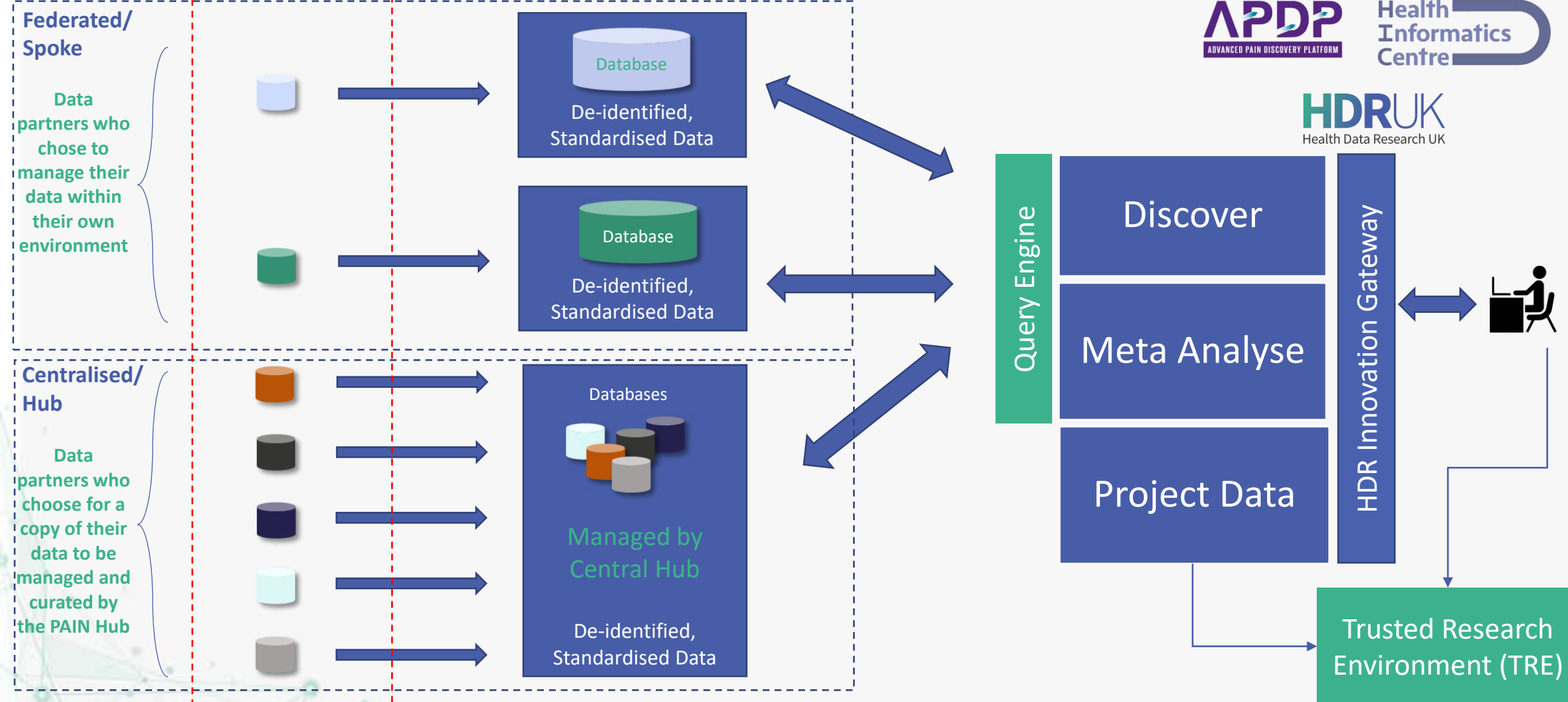
Alleviate
The Advanced Pain Discovery Platform
(APDP) Data Hub

HDRUK
Health Data Research UK

APDP
ADVANCED PAIN DISCOVERY PLATFORM

Health Informatics Centre

HDRUK
Health Data Research UK



Deliverables and Work Packages (WPs)

WP1: PPIE
• across all Hub WPs and wider APDP

WP2: Programme Management
• Communication; training; engagement

WP3: Data On-boarding
• Discoverable datasets; flexible: federated or centralised

WP4: Federation & Automation
• Customised software; automated loading and data feeds

WP5: Next generation TRE and Infrastructure
• Operational hub environment; TRE analytical platform;

01 **Milestone**
Hub Established
Q3 2021

02 **Milestone**
Service Delivery
Q1 2023

03 **Milestone**
Impact, Sustainability and
Scalability
Q1 2024

Sustainable Future
development

Public and Patient Involvement and Engagement

Alleviate
The Advanced Pain Discovery Platform
(APDP) Data Hub

HDRUK
Health Data Research UK

APDP
ADVANCED PAIN DISCOVERY PLATFORM

Health Informatics Centre

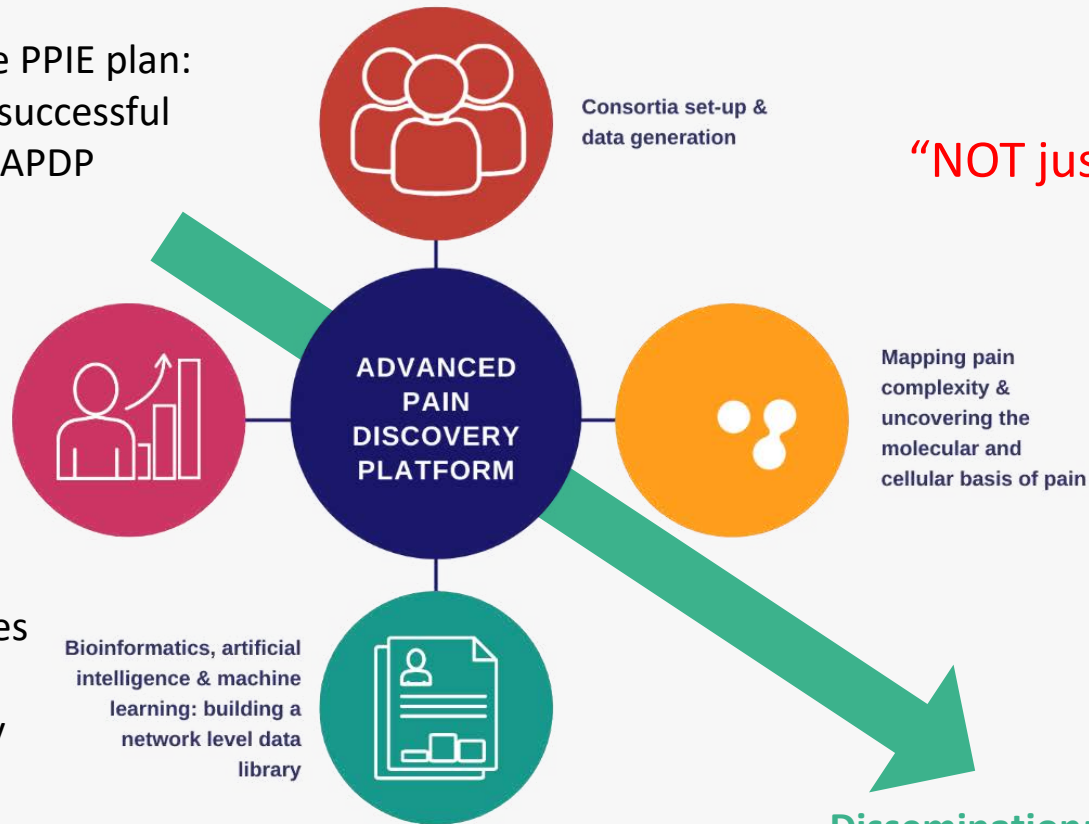
Co-create our cohesive PPIE plan:
working closely with successful
consortia across APDP

Iterative
process with
formal review

Online open
data platform

Data governance
Data on-boarding priorities
Research Priorities
Communication strategy

Day to day leadership from PPIE
representatives



“NOT just a tick box!”

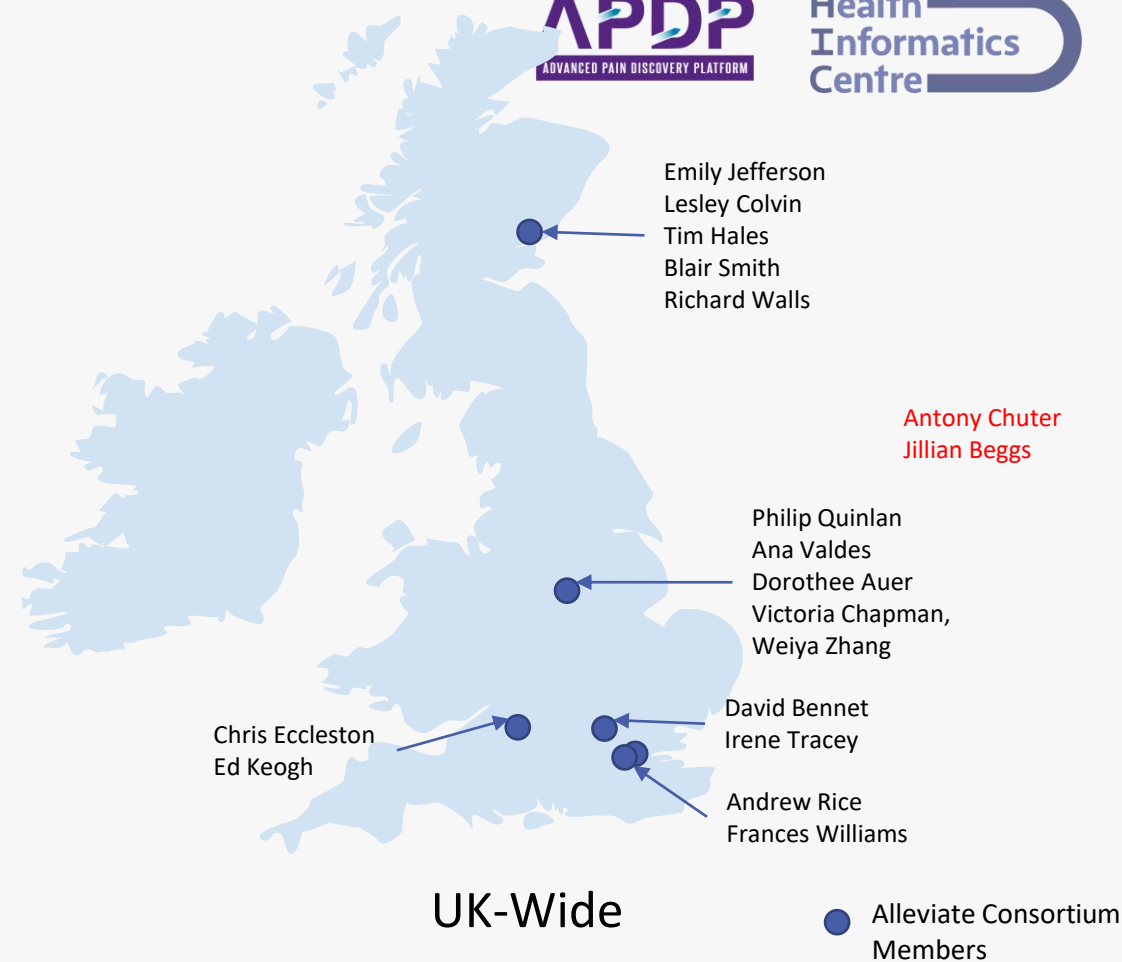
Jillian Beggs
Antony Chuter
Patient Insight Partners

PPIE Group
(6-8 PWLE)

Patient reference group
(~ 100 PWLE)

Dissemination:

People living with chronic pain,
carer, public, social and
healthcare professionals, third
sector, policy makers



Antony Chuter
Jillian Beggs

Summary of Datasets

Type	Total	Federated	Centralised
Large datasets of genomic and imaging data	6	4	2
Small, deep phenotyped datasets	16	9	7
Total	22	13	9

6 nodes in total
(as from 6 organisations)

- These are the datasets we have already committed to on-board
- There will be additional datasets from APDP consortia and **other groups**

*Central/federated choices will be confirmed during Alleviate i.e., proposal gave an indication of preference at this stage

Alleviate Summary

- Cohort discovery via the gateway – rather than just database descriptions and metadata fields
- Choice for data owners to share without needing to relinquish control
- Data harmonisation across data sources
- A Next Generation TRE – supporting multi-omic and imaging data along with AI
- Pain data covering different clinical areas along with population cohorts
- Importing pain phenotypes into the phenotype portal
- Strong desire to work with all other HDR hubs to share relevant data across clinical domains and shared learnings
- Research at pace and scale meaning it reaches patients faster
- Holistic approach to the patient experience focused on patient benefits rather than clinical specialties

Thanks for listening!

Happy to take questions....





DATA MIND

The Health Data Research Hub for Mental Health

DATAMIND OVERVIEW

Ann John and Robert Stewart

The leadership team



Ann John
PI and Co-director



Robert Stewart
Co-director



Andrew McIntosh
Chief Scientific Officer



Mathew Broadbent
Chief Operating Officer



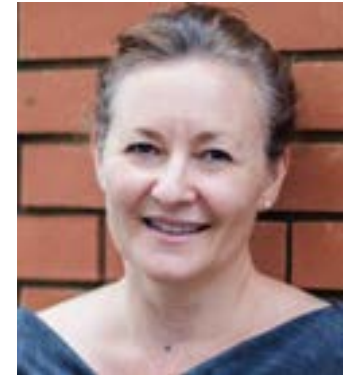
Dermott O'Reilly
Chief Data Officer



Lea Milligan
Core Activity Lead



Rudolf Cardinal
Core Activity Lead



Louise Arseneault
Core Activity Lead

Objectives



To **embed patient, personal experience and public participation** to ensure that the Hub is driven by the needs of the population and considers key ethical issues pertinent to MH data



Be **guided by and responsive to the needs of users** (Academia, the NHS, Third Sector, Policy Makers and Industry) in the data gathered and resources provided



To make datasets **visible, accessible and available** securely for research, development and innovation across Academia, the NHS, Third Sector, Policy Makers and Industry



To **curate and enhance the interoperability of data** for research, development and innovation across Academia, the NHS, Third Sector, Policy Makers and Industry



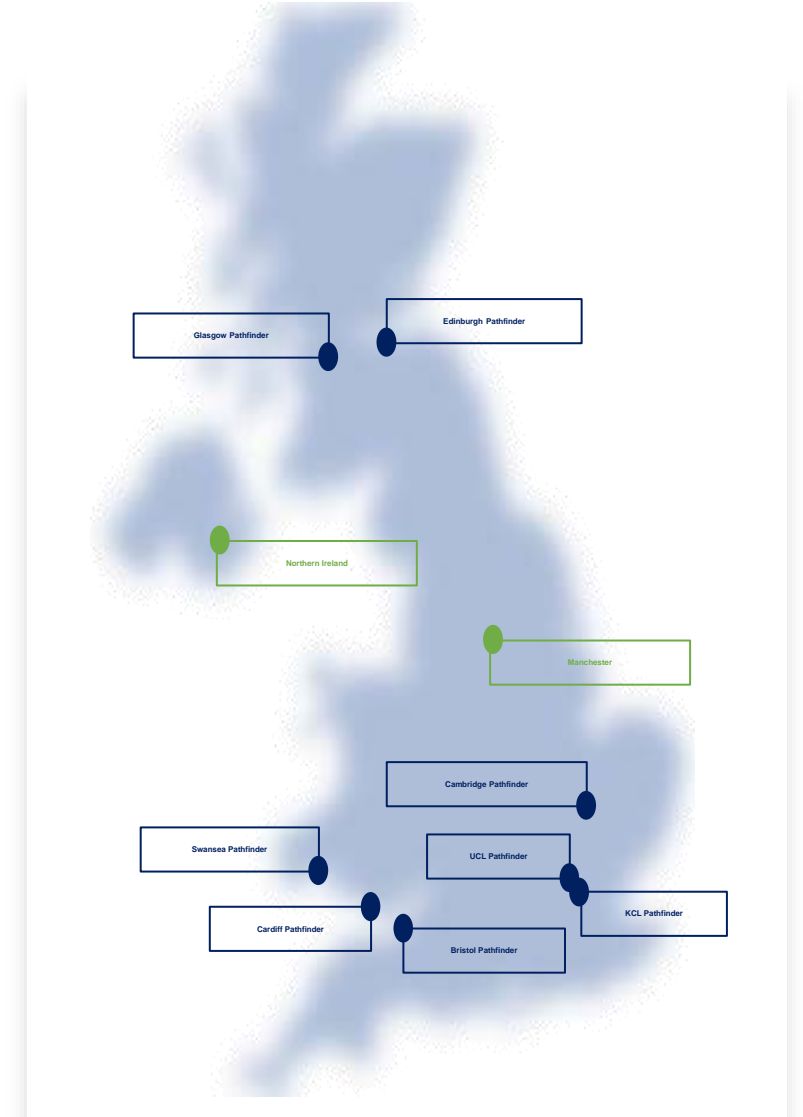
To **develop global standards** to meet industrial research and development needs



To **support capacity development**, including development of early career researchers

Our unique short-term benefits

- Benefits of **scale and reach – 4 nations**
- Building and amplifying **existing networks and activities**
- Building and sustaining **existing investments and expertise**
- Building and bridging **existing PPIE**
- Building **Capacity, MQ** at core
- Broad and multidisciplinary
- Brokering policy relevance
- Bringing National Centres
- Other Hubs



UNITING THE UK'S HEALTH DATA TO MAKE DISCOVERIES THAT IMPROVE PEOPLE'S LIVES

TRUSTWORTHY USE OF DATA



UK HEALTH DATA RESEARCH ALLIANCE

Uniting the UK's health data

UK Health Data Research Alliance

NHS DATA

COHORT DATA

REGISTRY DATA

TRIALS DATA

Convened by

HEALTH DATA RESEARCH HUBS

Curating disease focused datasets and providing
clinical trials and realworld evidence services



DATA MIND
The Health Data Research Hub for Mental Health



HEALTH DATA RESEARCH INNOVATION GATEWAY

Finding and accessing nationally aggregated, usable data

DISCOVER

DE-IDENTIFY

LINK

ACCESS

ANALYSE

PHENOTYPE LIBRARY

LINK TO TRUSTED RESEARCH
ENVIRONMENTS (SAFE HAVENS)

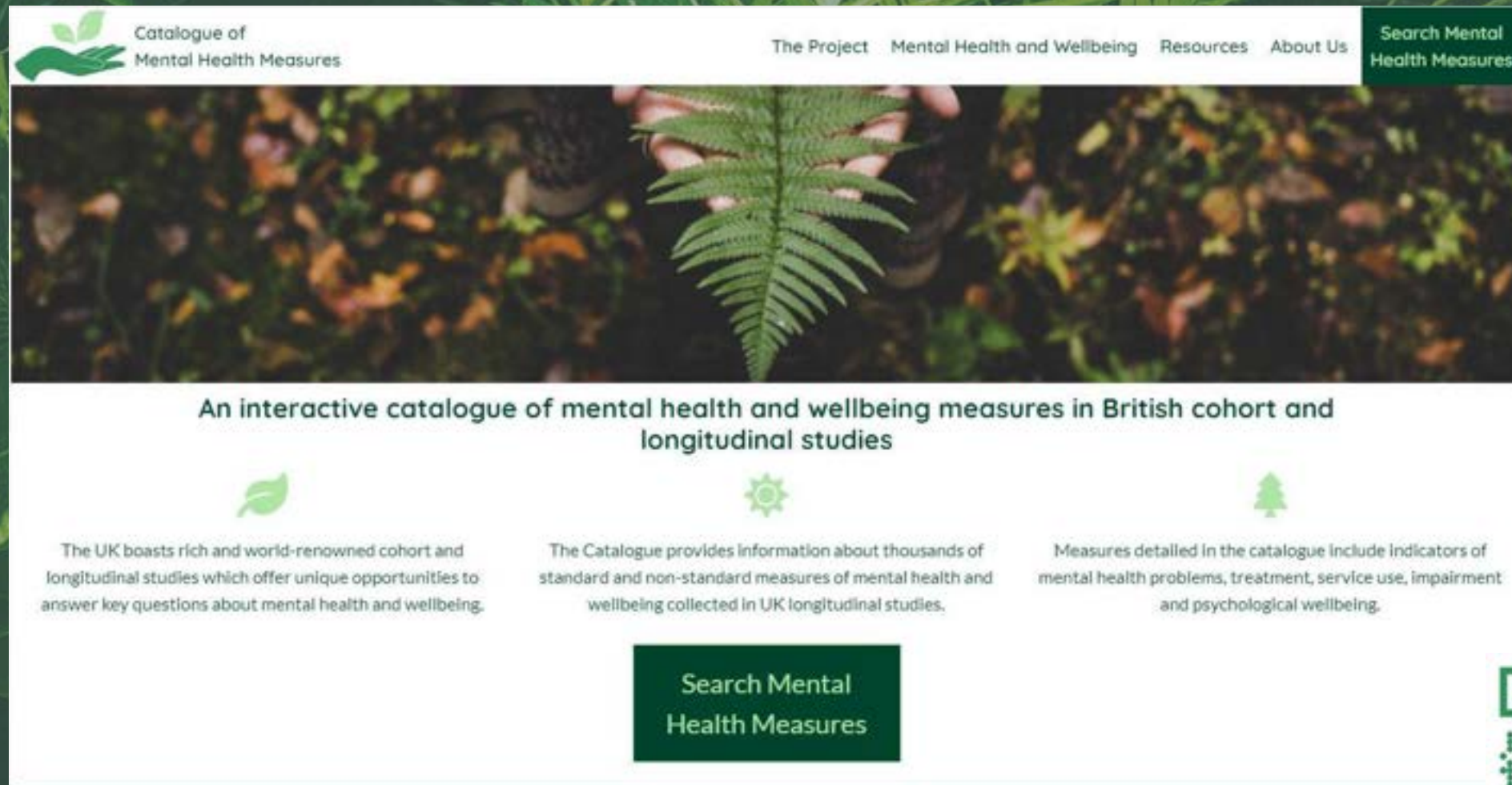
Operated by



USERS:
Researchers, Industry,
NHS, Charities and
Government

PATIENT AND PUBLIC BENEFIT

Catalogue of Mental Health Measures



The screenshot shows the homepage of the 'Catalogue of Mental Health Measures' website. The header features the site's logo (a green leaf icon) and navigation links: 'The Project', 'Mental Health and Wellbeing', 'Resources', 'About Us', and a 'Search Mental Health Measures' button. Below the header is a large image of a hand holding a green fern frond over a forest floor. The main content area has the title 'An interactive catalogue of mental health and wellbeing measures in British cohort and longitudinal studies'. It features three columns of text, each with a green icon (a leaf, a gear, and a tree respectively). At the bottom center is a large green button labeled 'Search Mental Health Measures'. In the bottom right corner, there is a QR code that links to the website.

Catalogue of Mental Health Measures

The Project Mental Health and Wellbeing Resources About Us Search Mental Health Measures


An interactive catalogue of mental health and wellbeing measures in British cohort and longitudinal studies

The UK boasts rich and world-renowned cohort and longitudinal studies which offer unique opportunities to answer key questions about mental health and wellbeing.

The Catalogue provides information about thousands of standard and non-standard measures of mental health and wellbeing collected in UK longitudinal studies.

Measures detailed in the catalogue include indicators of mental health problems, treatment, service use, impairment and psychological wellbeing.

Search Mental Health Measures



www.cataloguementalhealth.ac.uk/

– the rationale plans for public engagement within DATAMIND

A spectrum of sensitivity:

aggregated data on large numbers of people

|

anonymised individual data with consent (e.g. for public sharing)

|

(e.g. from specific research projects)

de-identified routine NHS patient-level data without consent

|

(of varying levels of detail)

identifiable NHS data without consent

(rare for research but possible)

- Legal and governance requirements for each differ.
- Note: DATAMIND aims to support a central index, but not to put all relevant data in a big “pot” – many data sources will remain under local information governance procedures, but be more discoverable, via the HDR UK Innovation Gateway.



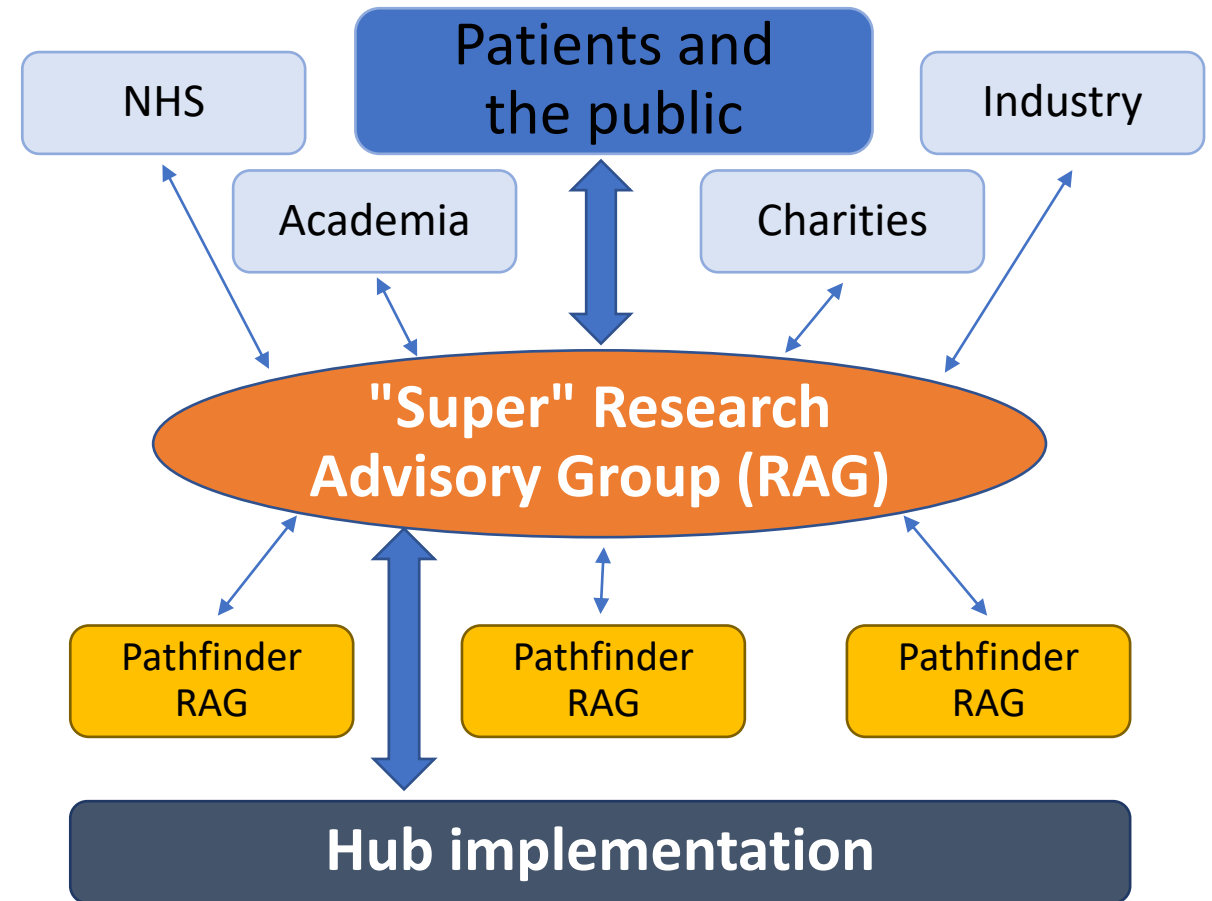
DATA MIND

The Health Data Research Hub for Mental Health

- **NHS Constitution for England (2013, 2015)**
- “The NHS... commits: ...
- to anonymise the information collected during the course of your treatment and use it to support research and improve care for others (pledge); ...
- to inform you of research studies in which you may be eligible to participate (pledge)”
- *A general principle:*
 - – “Consent or anonymise.”

Patient & Public Involvement and Engagement (PPIE) strategy

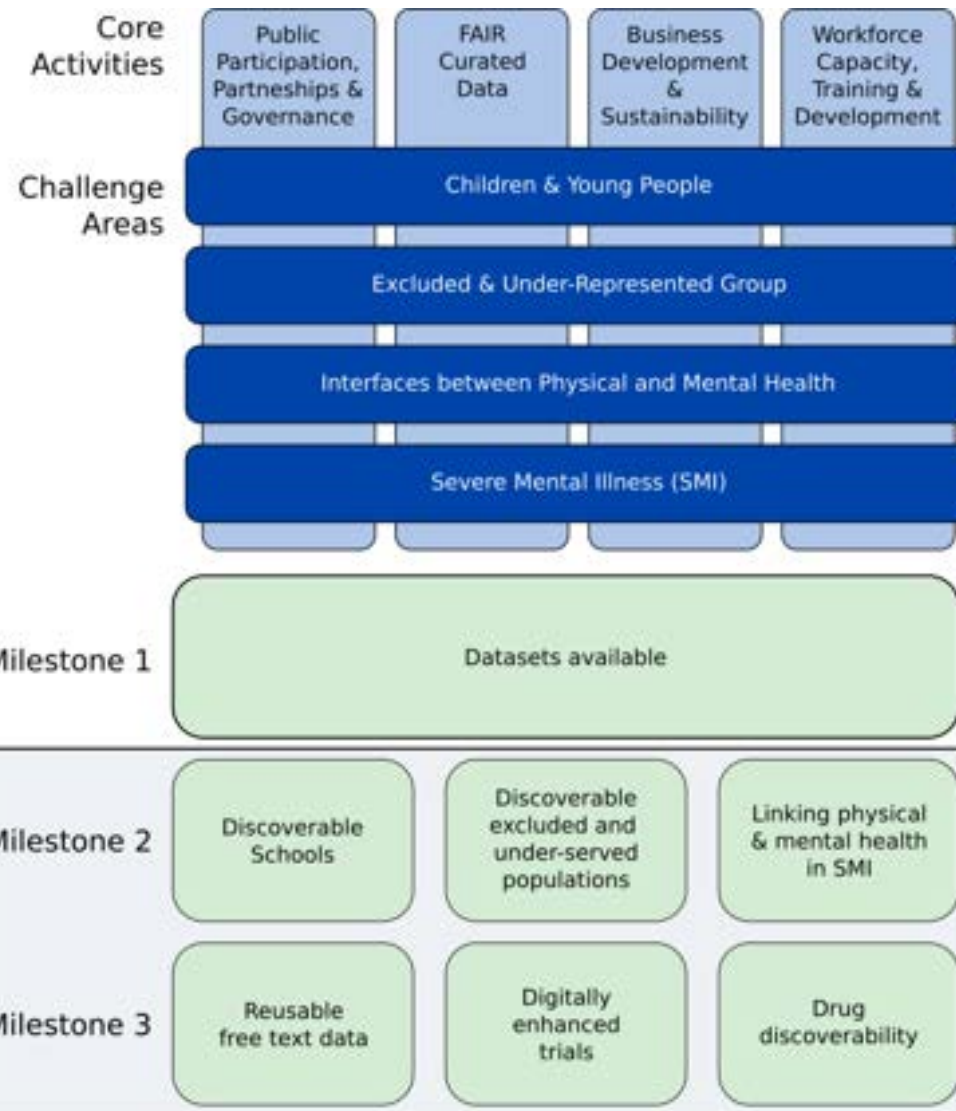
- Proposal shaped by PPI input and UK-wide survey
- Builds on existing PPIE investment
- PPIE embedded throughout
- Develop global standards
- Data literacy course





DATA MIND

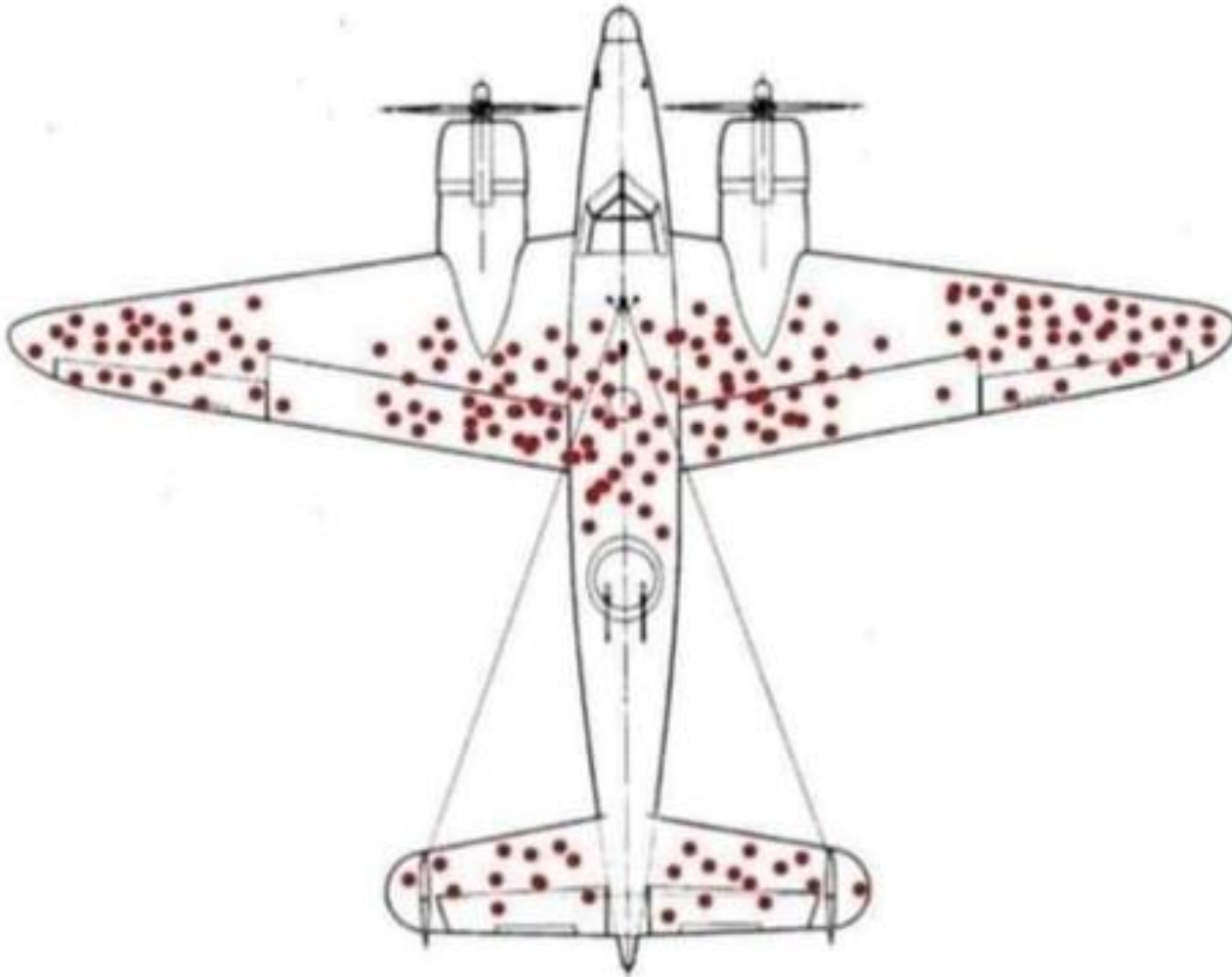
The Health Data Research Hub for Mental Health



DATAMIND and UK field- leadership

- Resources
 - Cohorts
 - Bioresource
 - EHR data
 - Administrative data (non-health)
 - Other data
 - Geospatial
 - Social media trends
 - Etc.
- Challenges
 - Harmonisation (without extra constraints)
 - Accessibility (in reality, beyond 'discoverable')
 - Linkages and networks
 - Data depth and visibility (NLP)
 - Local/National dichotomy

Using Healthcare Data for Biomedical Research



Healthcare data are diverse, incomplete, inaccurate, inconsistent, biased, misleading and poorly contextualised.

Analysis of such data may lead to high levels of 'noise' and erroneous interpretations.

Effect sizes may well be far stronger than those from real biomedical phenomena.

The danger of drawing false conclusions is therefore considerable.

This could be made worse when data are forced into limited standardised data models.

How real/common are such problems, are there defensive strategies and alternatives?

Provenance and Generalisability of Registry & EHR Data

Keith R Abrams

Department of Statistics, University of Warwick

17/03/2022

Acknowledgments: Jonathan Broomfield, Suzanne Freeman, Micki Hill,
Mark Rutherford, *University of Leicester* & Michael Crowther, Alessandro
Gasparini, *Karolinska Institutet*



Outline

- Issues surrounding use of **specialist registries** (especially in rare diseases)
- **Informative observations** in routine/EHR data

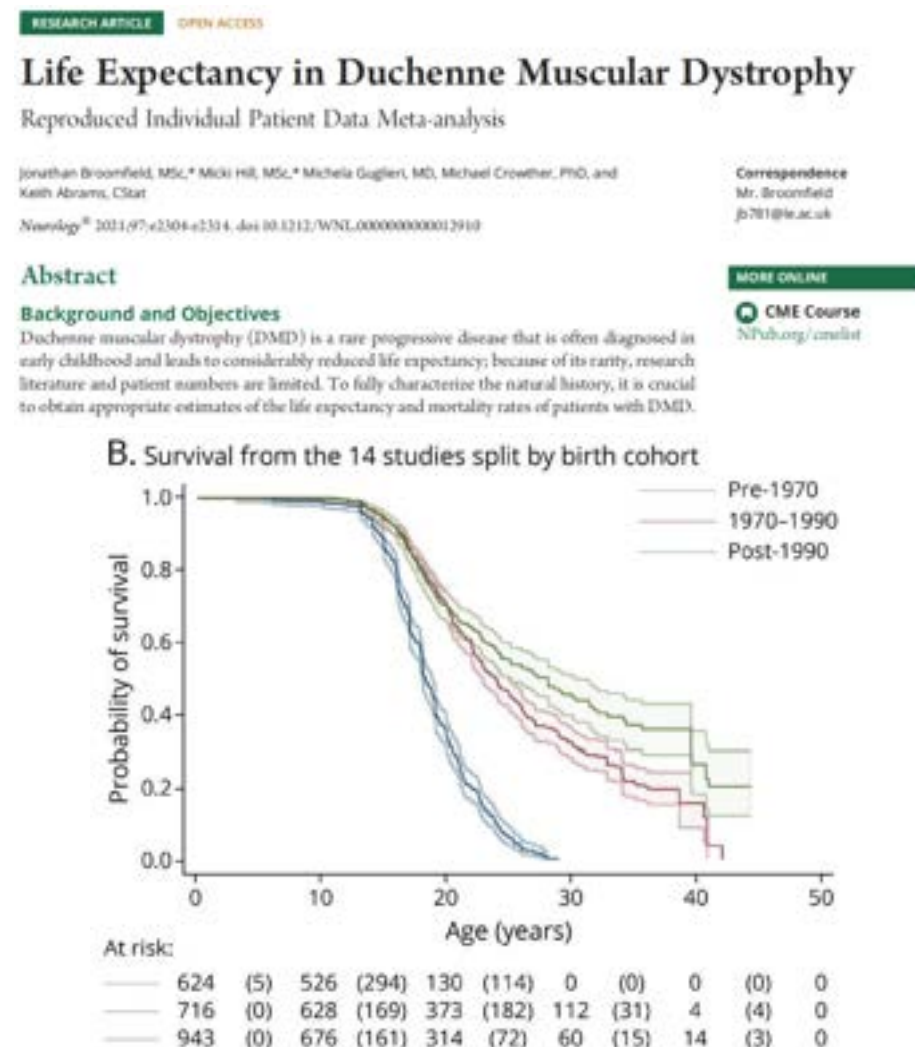
Use of Data from Specialist Registries

- In many diseases (but especially rare diseases) **specialist registries** are used to estimate patient and policy relevant quantities, e.g. life expectancy, Quality of Life etc.
- However, there is concern that due to the often **highly selected patient population** these registries may provide inappropriate estimates for decision making at a population level.
- Use of **population-based data sources** may provide more appropriate estimates (even if data granularity is not as great), e.g. CPRD, SAIL

Use of Data from Specialist Registries – Example

- To populate a Natural History Model (NHM) in **Duchenne Muscular Dystrophy (DMD)** estimates of life-expectancy were required
- Systematic review (and extraction/recreation of Individual Patient Data) of **published registry studies** was undertaken, but ...
- Senior clinical advisor at tertiary centre in US (and that published one of the studies used) remarked ... *“Some of our patients fly in for their appointments, and then they just stop coming ... never hear from them again.”*
- Currently using **CPRD** to undertake an analysis of UK patients (not based on their attendance at a specialist clinic, but rather population) to corroborate (or not) the estimates obtained from systematic review

Broomfield J, et al. Life Expectancy in Duchenne Muscular Dystrophy: Reproduced Individual Patient Data Meta-analysis. *Neurology*. 2021 Dec 7;97(23):e2304-e2314. doi: 10.1212/WNL.0000000000012910.



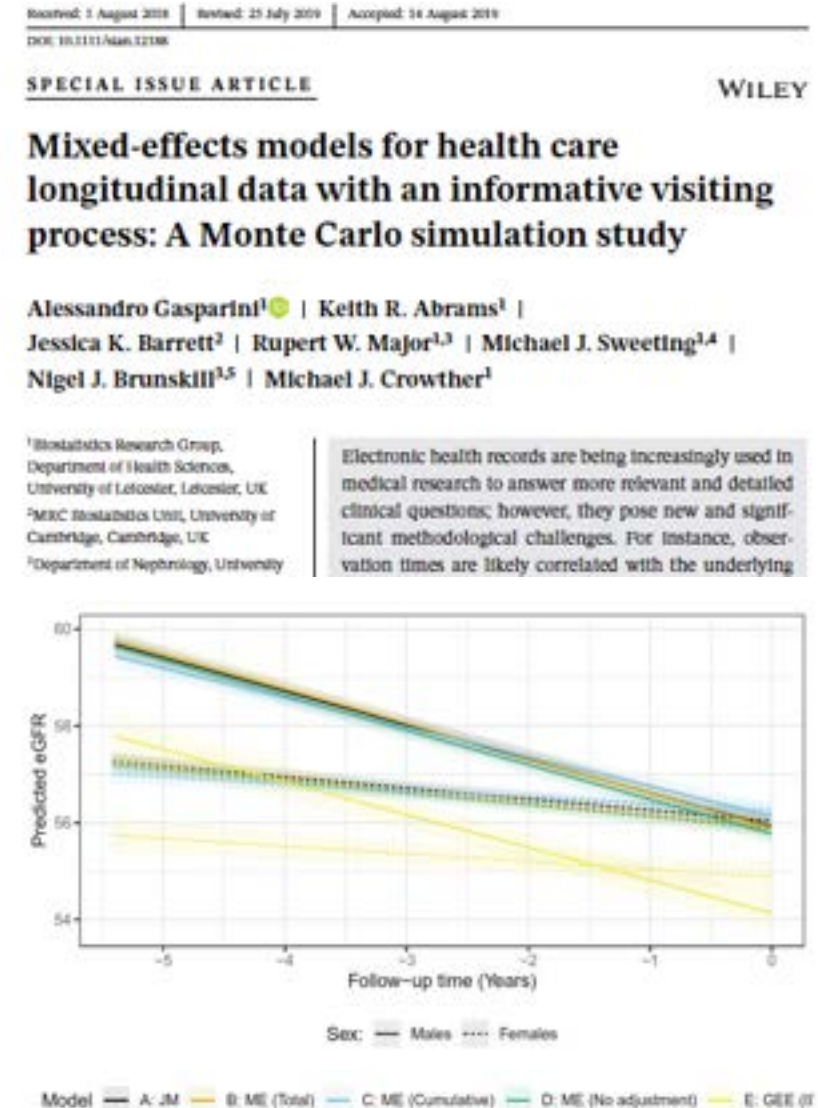
Informative observations in routine/EHR data

- Routine/EHR data provide rich source of information on patient pathways (even if not always with the granularity we might wish, e.g. Quality of Life)
- But in analysing such data, we need to consider **why the data (we have) are there?**
- For example, if we are interested in relating changing blood pressure to CVD, some patients may have many more bp readings than others (and for good reason, e.g. treatment), and **analysing all available bp readings** may mask any relationship at a population level.

Informative observations in routine/EHR data - Example

- Based on a trial in **Chronic Kidney Disease (CKD)** in which we were interested in **profile of eGFR** over time, undertook a simulation study of available methods;
- Approaches included;
 - Linear mixed model (**ignoring visiting process**)
 - Linear mixed model but **adjusting for number of visits**
 - Marginal model with **inverse intensity of visit weights**
 - Joint model – **model both eGFR & visitation process**
- Ignoring visitation process can lead to **biased results** and in CKD example different methods led to **different predictions** of eGFR profile over time

Gasparini A, et al. Mixed-effects models for health care longitudinal data with an informative visiting process: A Monte Carlo simulation study. *Stat Neerl*. 2020 Feb;74(1):5-23. doi: 10.1111/stan.12188.



Health Data Poverty

A barrier to inclusive and equitable AI

Prof Alastair Denniston



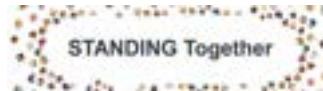
Oxford University Hospitals
NHS Foundation Trust



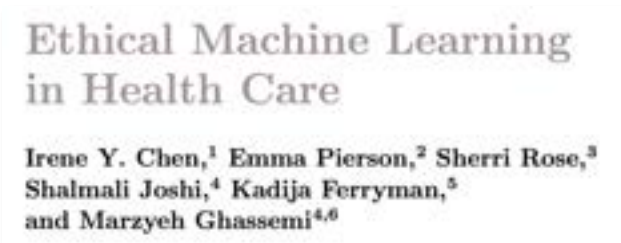
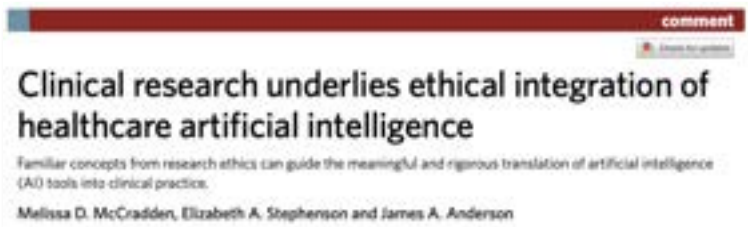
UNIVERSITY OF
BIRMINGHAM



University Hospitals Birmingham
NHS Foundation Trust



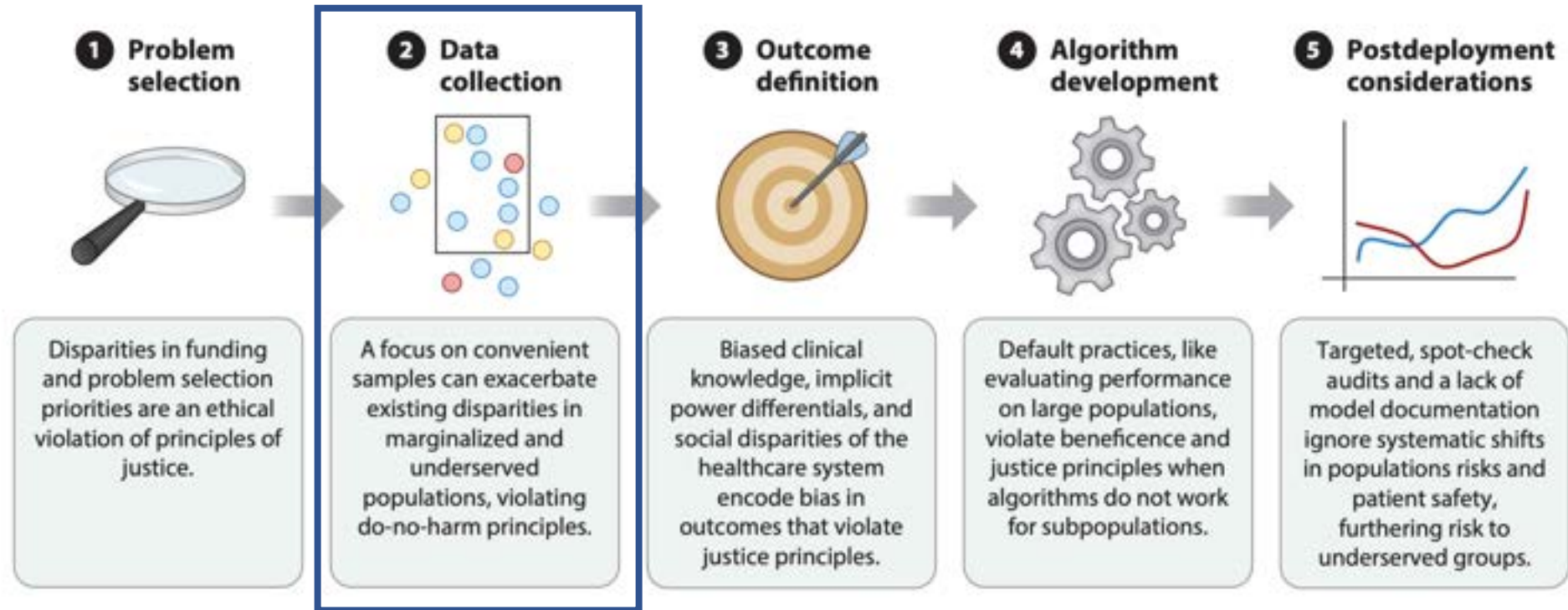
AI, bias and selective failure is an area of increasing concern



A major component of AI bias is the 'data foundation' on which AI is built

Annual Review of Biomedical Data Science
Ethical Machine Learning
in Healthcare

Irene Y. Chen,¹ Emma Pierson,² Sherri Rose,³
Shalmali Joshi,⁴ Kadija Ferryman,⁵
and Marzyeh Ghassemi^{1,6}



We and others have shown that these data foundations are poorly reported and lack diversity or representativeness

A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability

Saad M Khan*, Xiaoxuan Liu*, Subharthi Neethi, Edward Korost, Livia Fares, Siegfried K Wagnen, Peirce A Keane, Neil J Sebire, Matthew J Burton, Alastair K Denniston



Characteristics of publicly available skin cancer image datasets: a systematic review

David Wirt, Saad M Khan, Antonio Ji Xu, Hussain Ibrahim, Luke Smith, Jose Caballero, Luis Zepeda, Carlos de Blas Perez, Alastair K Denniston, Xiaoxuan Liu*, Roberto N Martin*

Support the Guardian
Available for everyone, funded by readers

Sign in The Guardian News website of the year

Subscribe → Contribute →

News Opinion Sport Culture Lifestyle

UK World Coronavirus Climate crisis Football Business Environment UK politics More

Skin cancer

AI skin cancer diagnoses risk being less accurate for dark skin - study

Research finds few image databases available to develop technology contain details on ethnicity or skin type

The Geographic Bias in Medical AI Tools

SHANA LYNN September 21, 2020

Home / Blog

Patient data from just three states trains most AI diagnostic tools.

SHARE THIS

Twitter Facebook Print



We and others have shown that these data foundations are poorly reported and lack diversity or representativeness

A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability

Saad M Khan*, Xiaoxuan Liu*, Subharthi Nath, Edward Korot, Livia Fries, Siegfried K Wagner, Peirce A Keane, Neil J Selvar, Matthew J Burton, Alastair K Denniston



Health data poverty: an assailable barrier to equitable digital health care

Hussein Ibrahim, Xiaoxuan Liu, Nevine Zariffa, Andrew D Morris*, Alastair K Denniston*

Health Data Poverty:

“The inability for individuals, groups, or populations to benefit from a discovery or innovation due to insufficient data that are adequately representative of them”

STANDING Together: collaboration funded by AI Ethics Award launched Oct 2021



Dr Xiao Liu
Clinical Research Fellow
University Hospitals
Birmingham NHSFT



Prof Alastair Denniston
Director of INSIGHT
University Hospitals
Birmingham NHSFT



[News](#)
[Opinion](#)
[Sport](#)
[Culture](#)
[Lifestyle](#)
[More](#)

[UK](#)
[World](#)
[Coronavirus](#)
[Climate crisis](#)
[Football](#)
[Business](#)
[Environment](#)
[UK politics](#)
[Education](#)
[Society](#)
[Science](#)
[Tech](#)

Artificial intelligence (AI)

AI projects to tackle racial inequality in UK healthcare, says Javid

Exclusive: health secretary signs up to hi-tech schemes countering health disparities and reflecting minority ethnic groups' data

Andrew Gregory

Wed 20 Oct 2021 06:00 BST

[f](#)
[t](#)
[e](#)

▲ AI robot, specialised for traditional Chinese medicine, shown in Beijing, 2020. In the UK, the government hopes new AI technology will lead to better healthcare training. Photograph: Xinhua/Rex/Shutterstock

Artificial intelligence is to be used to tackle racial inequalities in the NHS under government plans to “level up” healthcare.

STANDING Together: collaboration funded by AI Ethics Award launched Oct 2021



Dr Xiao Liu
Clinical Research Fellow
University Hospitals
Birmingham NHSFT



Prof Alastair Denniston
Director of INSIGHT
University Hospitals
Birmingham NHSFT



We believe health datasets should be curated with inclusivity and diversity in mind. To ensure AI healthcare technologies are supported by adequately representative data, we are developing standards on how AI datasets should be composed (*'who' is represented in the data*) and transparency around the data composition (*'how' they are represented*).

1

Defining essential characteristics for datasets through multistakeholder consensus

2

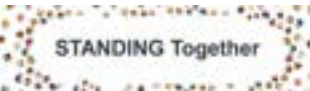
Dataset mapping through systematic review of datasets in cancer, cardiovascular disease & COVID-19

3

Identifying and Barriers through structured interviews with data curators

Thank you

a.denniston@bham.ac.uk



‘Trustworthiness’ in PROMs data

Dr Lee Aiyegbusi

*Research fellow, Deputy Director, Center for Patent
Reported Outcomes Research (CPROR)*

17/03/2022



Disclosures

- Supported by: National Institute of Health Research (NIHR) Birmingham Biomedical Research Centre (BRC), NIHR Applied Research Centre (ARC), West Midlands at the University of Birmingham and University Hospitals Birmingham NHS Foundation, Innovate UK (part of UK Research and Innovation), Gilead Sciences Ltd, and Janssen Pharmaceuticals, Inc.
- Personal fees: Gilead Sciences Ltd, GlaxoSmithKline and Merck.

Objectives



**BRIEFLY INTRODUCE PATIENT-
REPORTED OUTCOME MEASURES
(PROMS)**



**HIGHLIGHT CONSIDERATIONS FOR
PROM DESIGN**



**HIGHLIGHT CONSIDERATIONS FOR
PRO DATA COLLECTION**

Introduction to PROs



What Are Patient-reported Outcomes (PROs)?



A PRO is any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else



How Are PROs Captured?

- PROs may be captured using **self-reported** validated measures or questionnaires known as patient-reported outcome measures (PROMs)
- PROMs measure patients' health status
- Differ from patient-reported experience measures (PREMs), which focus on processes of care



Potential value and use of PROMs in health care



Measure patient priorities (FDA 2017)



Shared decision-making (Hutchings & Alrubaiy 2017)



Tailor care to patient need (Calvert et al. 2019)



Improved outcomes (Basch et al. 2015)



Safety endpoints in clinical trials (FDA 2009, 2020)

CONSIDERATIONS FOR PROM design



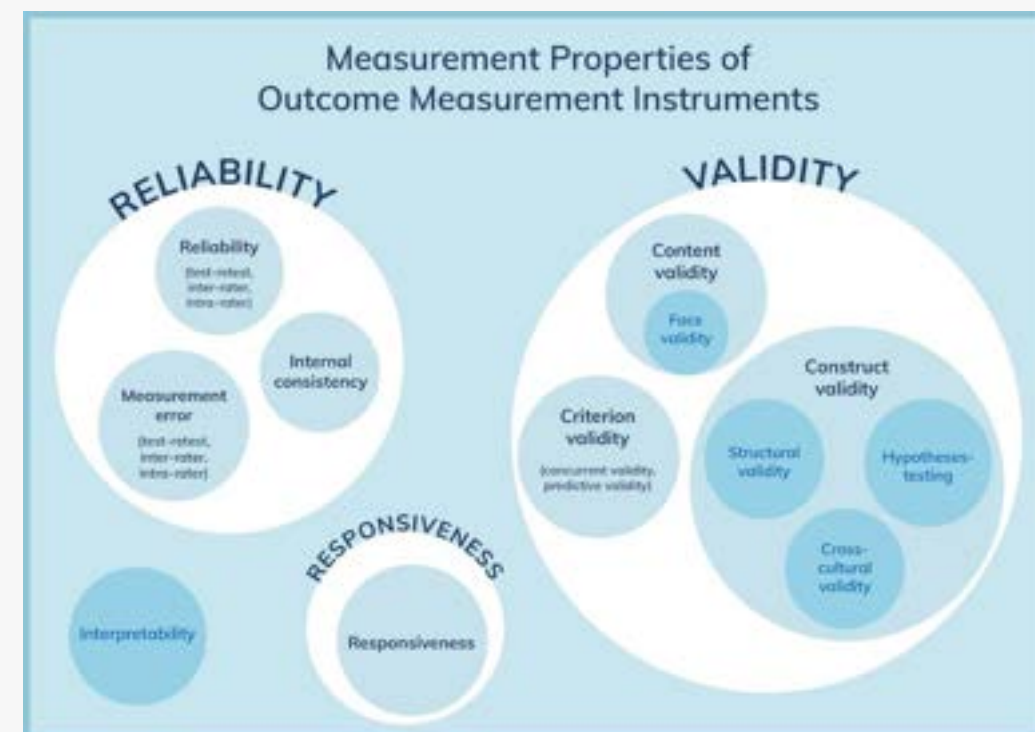
Considerations for PROM design (1)

- Stakeholder involvement
 - Item generation and selection should involve patients and clinicians to ensure comprehensiveness and relevance
 - Cognitive testing needs to be conducted with patients to ensure comprehension



Considerations for PROM design (2)

- The psychometric properties of the PROM (i.e., its qualities or attributes)
- The reliability, validity and responsiveness of measures should be determined
- Content validity is considered the most important property

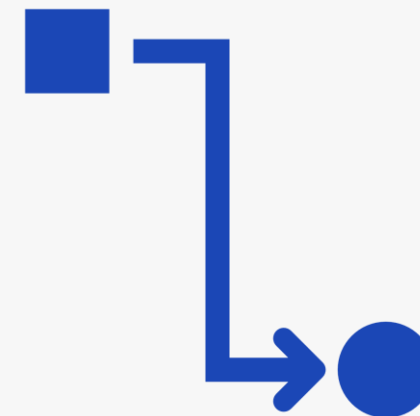


Considerations for PRO data collection



Considerations for PRO data collection (1)

- Selection of suitable PROM for the purpose and target population
 - Ideally the measure should be validated in target population
- Self-completion preferred
 - Proxy reporting may be biased
- Appropriately translated versions should be employed
 - Cross-cultural validity should be established



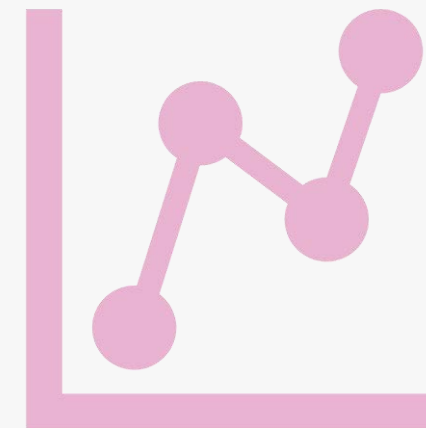
Considerations for PRO data collection (2)

- Mode of collection
 - Paper vs electronic format
 - Measurement equivalence has been established however, ePRO design issues may affect the data collected
 - ePRO collection removes human administrative error
 - Usability testing of ePRO systems with patients essential
 - Some patients still prefer paper so this option should be provided



Considerations for PRO data collection (3)

- Missing data can have major impacts on statistical analysis data.
 - Can produce wider confidence intervals and reduced statistical power to detect a treatment effect
 - Could distort treatment effect
- Efforts must be made to minimise the level of missing data
 - Treat PROs like any other outcome by integrating into the data collection and clinical trial protocol
 - Text reminders
 - Prompts when questions are skipped for ePROs
 - Paper versions checked for completion



Sloan JA et al. Analysis and interpretation of results based on patient-reported outcomes. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2007;10 Suppl 2:S106-15.

Calvert M et al. Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. Jama. 2018;319(5):483-94.

Calvert M, King M, Mercieca-Bebber R, Aiyegbusi O, Kyte D, Slade A, et al. SPIRIT-PRO Extension explanation and elaboration: guidelines for inclusion of patient-reported outcomes in protocols of clinical trials. BMJ Open. 2021;11(6):e045105.

Conclusion

There are numerous benefits that may be derived from using PROs in clinical trials and routine clinical practice.

However, to be truly beneficial, efforts should be made to ensure the data collected are trustworthy.

References

- FDA Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009.
- Aiyegbusi OL, Kyte D, Cockwell P, et al. A patient-centred approach to measuring quality in kidney care: patient-reported outcome measures and patient-reported experience measures. *Curr Opin Nephrol Hypertens*. 2017;26:442–449.
- Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights* 2013;6:61–68.
- Calvert M, Kyte D, Price G, et al. Maximising the impact of patient reported outcome assessment for patients and society. *BMJ (Clinical Research ed)* 2019;364.
- Hutchings HA, Alrubaiy L. Patient-Reported Outcome Measures in Routine Clinical Care: The PROMise of a Better Future? *Dig Dis Sci*. 2017 Aug;62(8):1841-1843.
- Basch E, Deal AM, Kris MG, et al. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. *Journal of Clinical Oncology* 2016; 34: 557-565.
- Mokkink et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research* 2018; 27: 1171-1179.
- Aiyegbusi OL, Nair D, Peipert JD, Schick-Makaroff K, Mucsi I. A narrative review of current evidence supporting the implementation of electronic patient-reported outcome measures in the management of chronic diseases. *Therapeutic Advances in Chronic Disease*. 2021.
- Snyder CF, Watson ME, Jackson JD, Cella D, Halyard MY. Patient-reported outcome instrument selection: designing a measurement strategy. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2007;10 Suppl 2:S76-85.
- Devlin N and Appleby J. Getting the most out of PROMs: Putting health outcomes at the heart of NHS decision-making. London: Kings Fund and Office of Health Economics 2010.
- LeRouge C, Austin E, Lee J, et al. ePROs in Clinical Care: Guidelines and Tools for Health Systems. Seattle, WA: CERTAIN, University of Washington, 2020.
- Aiyegbusi OL, Kyte D, Cockwell P, Marshall T, Dutton M, Walmsley-Allen N, et al. Development and usability testing of an electronic patient-reported outcome measure (ePROM) system for patients with advanced chronic kidney disease. *Computers in Biology and Medicine*. 2018;101:120-7.
- Sloan JA et al. Analysis and interpretation of results based on patient-reported outcomes. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2007;10 Suppl 2:S106-15.
- Calvert M et al. Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. *Jama*. 2018;319(5):483-94.
- Calvert M, King M, Mercieca-Bebber R, Aiyegbusi O, Kyte D, Slade A, et al. SPIRIT-PRO Extension explanation and elaboration: guidelines for inclusion of patient-reported outcomes in protocols of clinical trials. *BMJ Open*. 2021;11(6):e045105.

Dr Lee Aiyegbusi

- **Email:** O.L.Aiyegbusi@bham.ac.uk
- **Find out more:**
 - CPROR: <https://www.birmingham.ac.uk/research/cpror/index.aspx>
 - CPROR: www.birmingham.ac.uk/prolearn

HDRUK
Health Data Research UK



Exploring the impact of data quality and completeness on international cancer survival differences

Better Care Insight Sharing Day – 24/02/2022

17/03/2022

Paul C Lambert



UNIVERSITY OF
LEICESTER

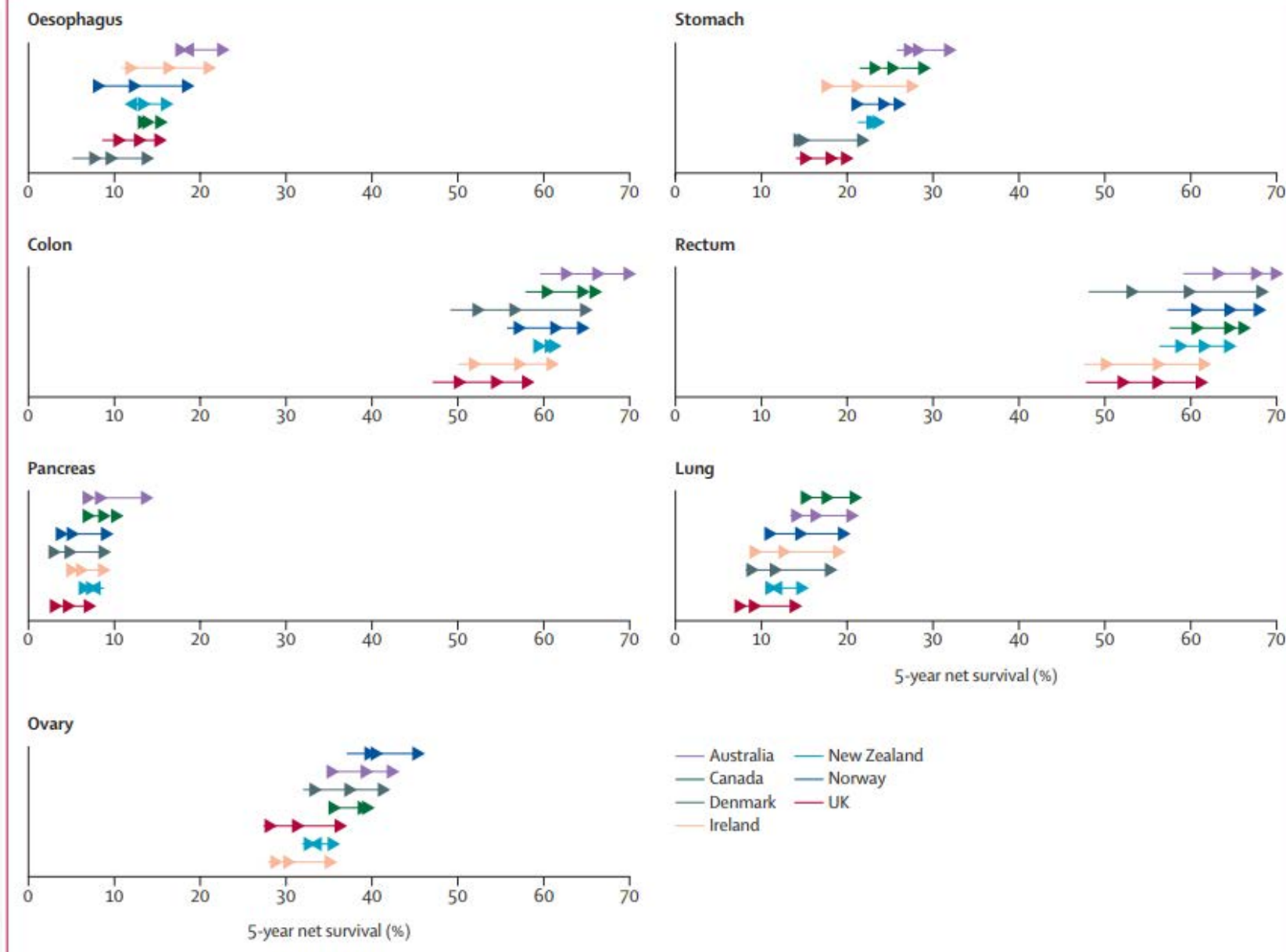


**Karolinska
Institutet**



Comparing Cancer Survival Between Countries

- There have been many comparisons of cancer survival between countries / regions. (e.g. EUROCARE I-VI, CONCORD, NORDCAN, ICBP)
- Valid comparisons require data that are collected, processed and coded in a comparable way.
- Survival comparisons are often criticized in that differences in survival could be due to differences in registration practice or completeness.
- I will describe some recent work as part of the International Cancer Benchmarking Project (<https://gco.iarc.fr/survival/survmark/>).



M Arnold et al. Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (ICBP SURVMARK-2): a population-based study. *Lancet Oncology* 2019.

Are these differences real or could they be due to differences in data quality / registry practice?

Outputs from ICBP (Element C)

ARTICLE
Epidemiology

Exploring the impact of cancer registry completeness on international cancer survival differences: a simulation study

Therese M.-L. Andersson^{a, *}, Mark J. Rutherford^{c, 2, 3}, Tor Åge Myklebust^{c, 5}, Bjørn Møller^d, Isabelle Soerjomataram³, Melina Arnold³, Freddie Bray³, D. Max Parkin^{3, 6}, Peter Sasieni⁷, Oliver Bucher⁸, Prithwish De⁹, Gerda Engholm¹⁰, Anna Gavin¹¹, Alana Little¹², Geoff Porter¹³, Agnihotram V. Ramanakumar¹⁴, Nathalie Saint-Jacques¹⁵, Paul M. Walsh¹⁶, Ryan R. Woods¹⁷ and Paul C. Lambert^{1, 2}



Contents lists available at ScienceDirect

Cancer Epidemiology

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

A way to explore the existence of “immortals” in cancer registry data – An illustration using data from ICBP SURVMARK-2

Therese M.-L. Andersson^{a, *}, Mark J. Rutherford^{b, c}, Tor Åge Myklebust^{d, e}, Bjørn Møller^d, Melina Arnold^c, Isabelle Soerjomataram^c, Freddie Bray^c, Hazem Abd Elkader^f, Gerda Engholm^g, Dyfed Huws^{h, i}, Alana Little^j, Lorraine Shack^{k, l}, Paul M. Walsh^m, Ryan R. Woodsⁿ, D. Maxwell Parkin^{o, p}, Paul C. Lambert^{a, b}

The impact of excluding or including Death Certificate Initiated (DCI) cases on estimated cancer survival: A simulation study

Therese M.-L. Andersson^{a, *}, Tor Åge Myklebust^{b, c}, Mark J. Rutherford^{d, e}, Bjørn Møller^b, Isabelle Soerjomataram^c, Melina Arnold^c, Freddie Bray^c, D. Max Parkin^{c, f}, Peter Sasieni^g, Oliver Bucher^h, Prithwish Deⁱ, Gerda Engholm^j, Anna Gavin^k, Alana Little^l, Geoff Porter^m, Agnihotram V. Ramanakumarⁿ, Nathalie Saint-Jacques^o, Paul M. Walsh^p, Ryan R. Woods^q, Paul C. Lambert^{a, d}

Can different definitions of date of cancer incidence explain observed international variation in cancer survival? An ICBP SURVMARK-2 study

Tor Åge Myklebust^{a, b, *,}, Therese Andersson^c, Aude Bardot^d, Sally Vernon^e, Anna Gavin^f, Deirdre Fitzpatrick^f, Marianne Brenn Jerm^a, Mark Rutherford^{d, g}, D. Maxwell Parkin^h, Peter Sasieniⁱ, Melina Arnold^d, Isabelle Soerjomataram^d, Freddie Bray^d, Paul C. Lambert^{c, g}, Bjørn Møller^a

British Journal of Cancer

www.nature.com/bjc

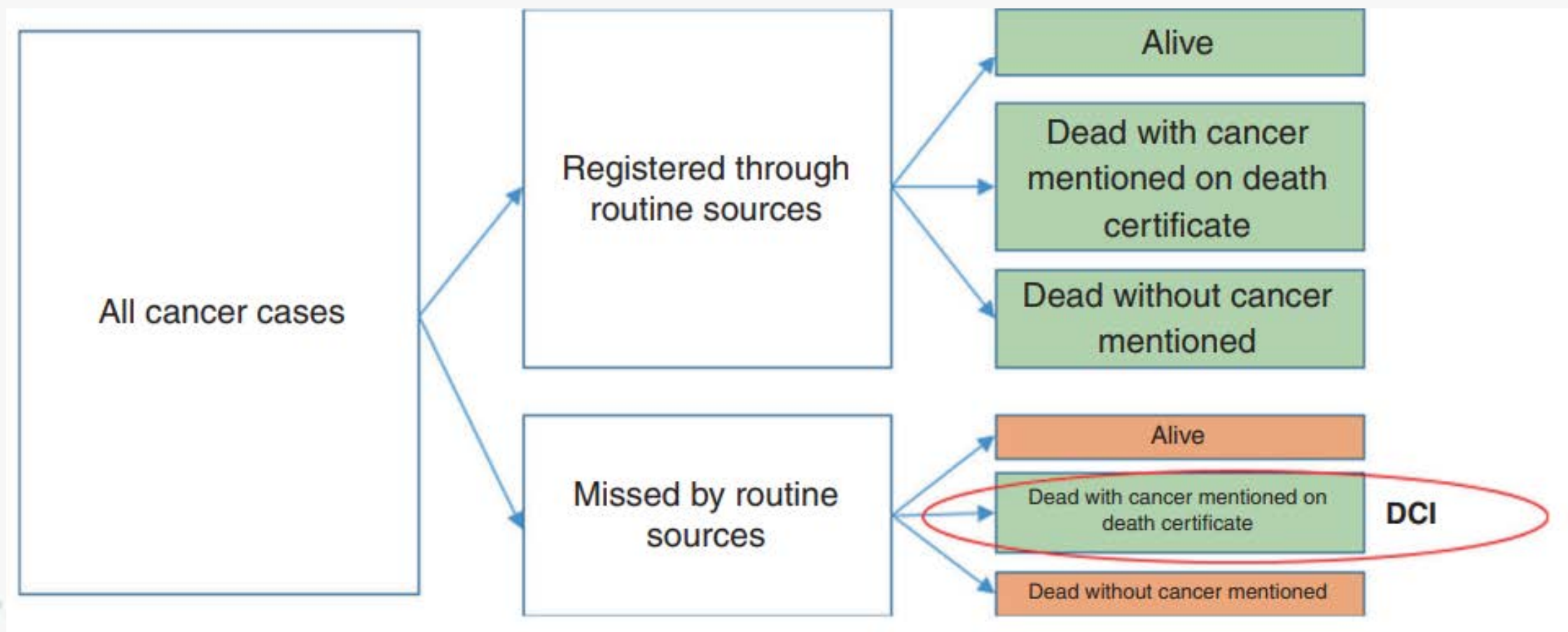
ARTICLE **OPEN**

Epidemiology

Five ways to improve international comparisons of cancer survival: lessons learned from ICBP SURVMARK-2

Therese M.-L. Andersson^{1, 5, *}, Tor Åge Myklebust^{2, 3}, Mark J. Rutherford^{4, 5}, Bjørn Møller², Melina Arnold⁵, Isabelle Soerjomataram⁵, Freddie Bray⁵, D. Maxwell Parkin^{6, 7} and Paul C. Lambert^{1, 4}

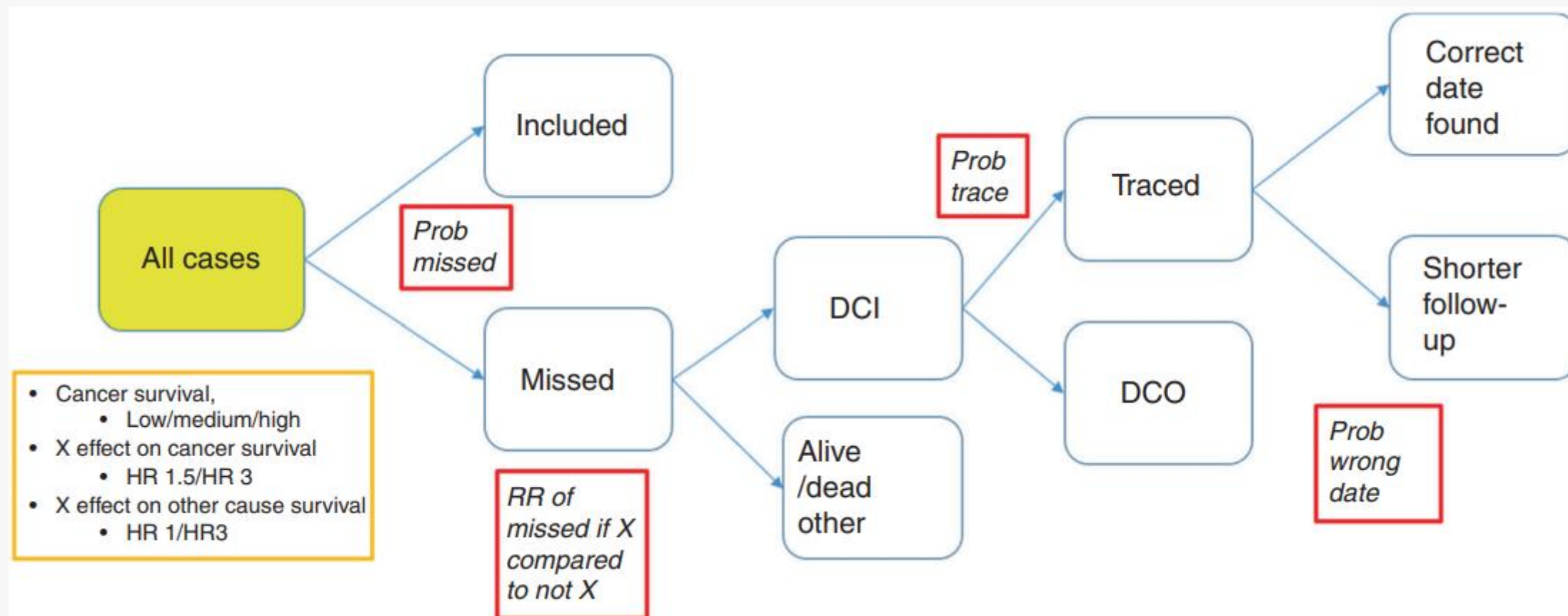
Using Death Certificates to Obtain Missed Cases (Traceback)



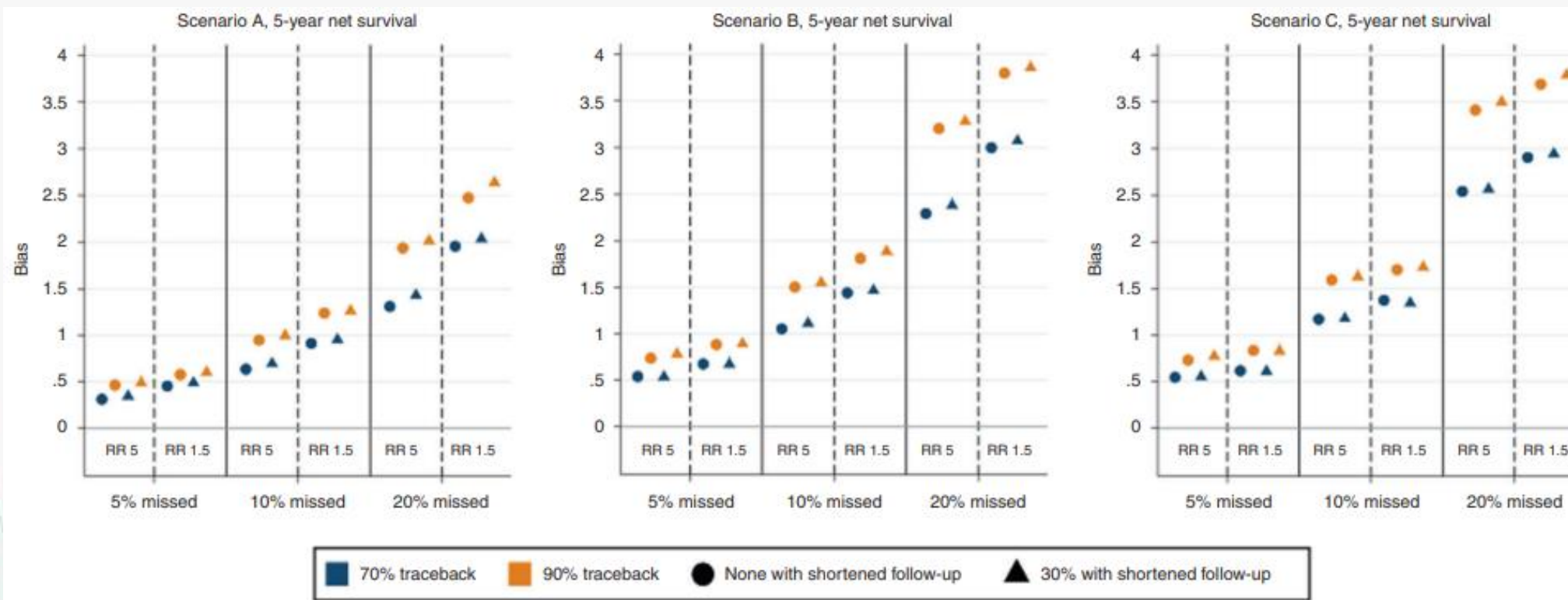
Simulating the “Perfect” Cancer Registry

1. We used simulation studies to understand how differences/imperfections in the registration can impact comparisons between countries.
2. We can simulate the “**perfect**” cancer registry.
 - When we simulate the “truth” is known.
3. We can then introduce various differences/errors.
 - a) missed individuals (and differences in prognosis)
 - b) incorrect dates of diagnosis.
 - c) The use (and intensity) of traceback to obtain missed individuals.
 - d) etc etc.
4. Can look at factors in isolation and in combination and quantify bias.
5. Repeat many (1000s) of times for various scenarios.
 - Useful to include “extreme” scenarios.

Andersson et al, British Journal of Cancer 2021; ;124:1026–32



Quantifying Bias Over Various Scenarios



Summary

- We should acknowledge that our datasets will never be perfect.
- We should understand how these imperfections affect our analyses and their interpretation.
- This will always depend on the type of analysis being performed.
- Creating simulated data can help us understand how imperfect data may impact on our real analyses.

Acknowledgements

Therese Andersson, Bjørn Møller, Mark J Rutherford, Tor Åge Myklebust, Isabelle Soerjomataram, Freddie Bray and the SURVMARK-2 team

The Challenge of Generating Realistic Synthetic Healthcare Datasets

Theodoros N. Arvanitis, RT, DPhil, CEng, FRSM, FGBHI

Associate Director, HDR UK Midlands

Professor of Digital Health Innovation & Director, Institute of Digital Healthcare, WMG, University of Warwick

Honorary Professor and Clinical Scientist, University Hospitals Coventry and Warwickshire NHS Trust

24 February 2022

Better Care Insight Sharing Day



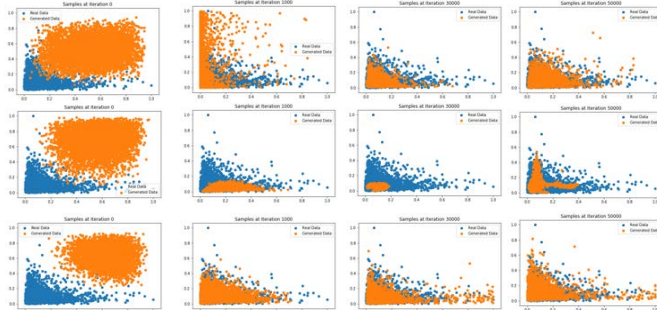
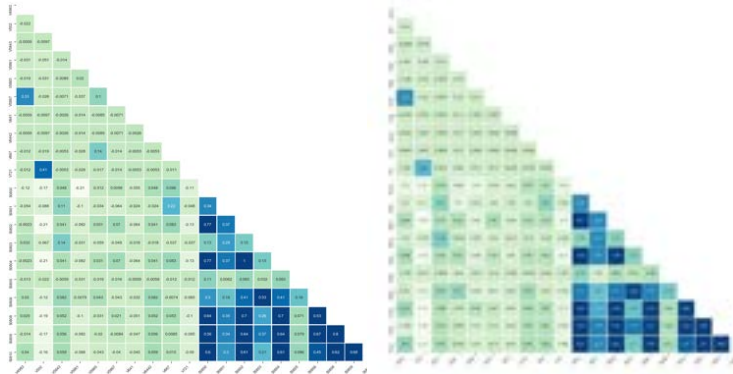
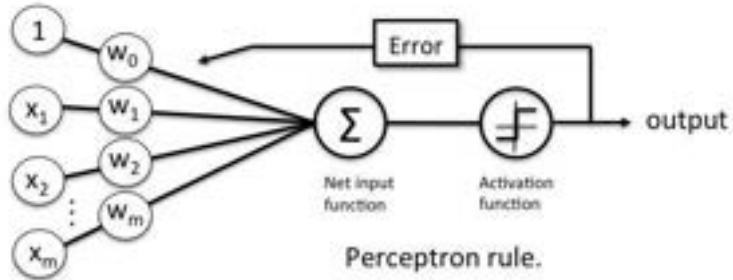
The context: facilitating safe and clinical valid digital health innovations

- Significant advances in digital health and health data science across numerous aspects
 - Diagnostic, self-monitoring, CDSS, telehealth
- Challenges in using health data sets in research and innovation
- Concerns
 - Missing Data, Data Representation, Data Bias
 - Technical safety
 - Clinical validation
- Researchers and innovators need access to data to validate applications
 - In conflict with privacy (and potentially confidentiality) concerns
- May deprive patients from new advances and innovations



Source: The Guardian <https://www.theguardian.com/technology/2017/jul/03/google-deepmind-16m-patient-royal-free-deal-data-protection-act>

Assurance of intelligent applications using realistic synthetic datasets



Ambition/Impact

Facilitate safe and clinically valid digital health innovation

Challenge

Privacy concerns restrict access to data needed for clinical validation and safety assurance, depriving of potential innovations

IDH Innovation

Generation of realistic synthetic datasets, equivalent to real ones, using advanced machine learning methods.

IDH Innovation

Identification of assurance framework and certification basis for innovations based on synthetic data.

Staff

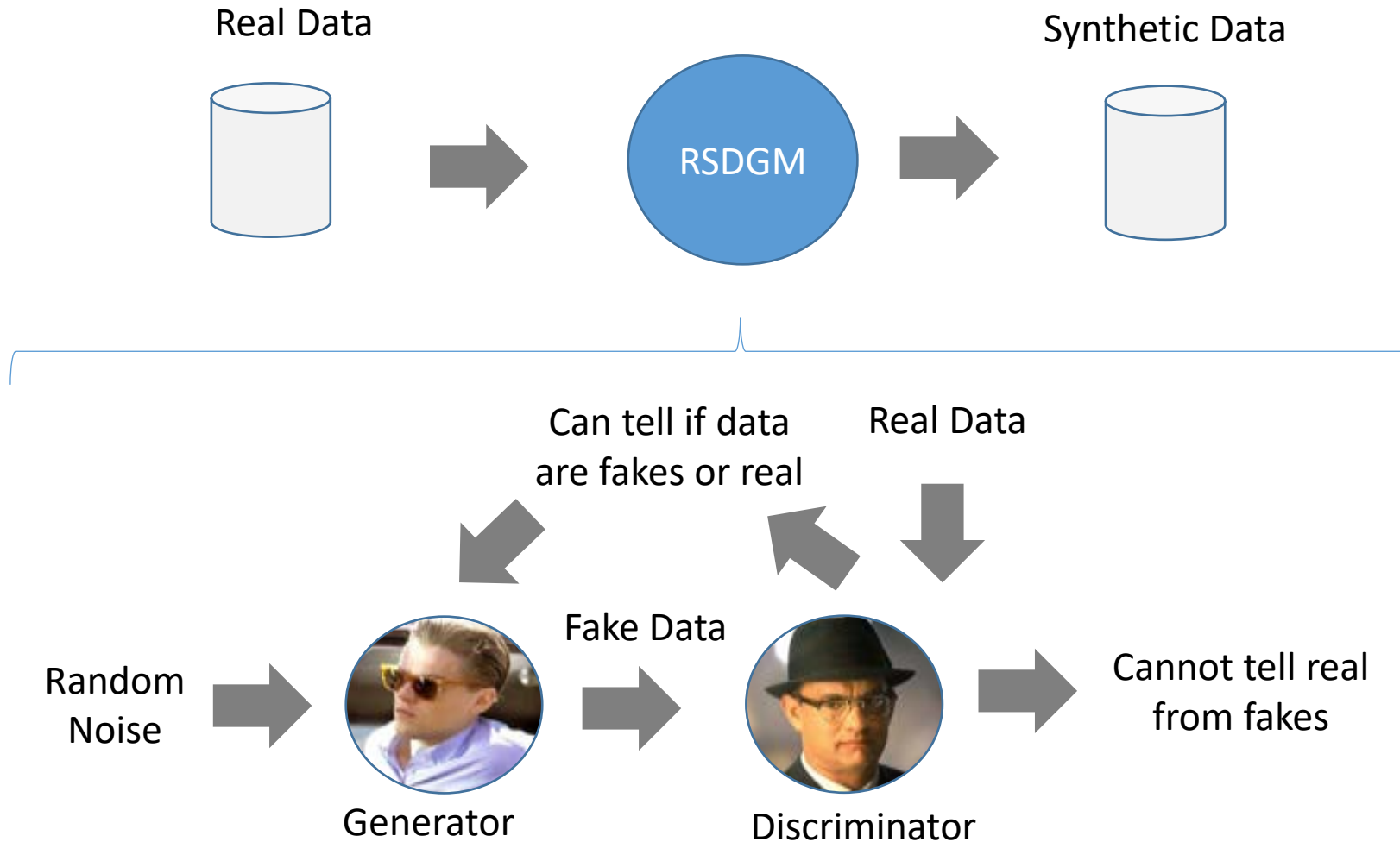
George Despotou, Eda Ozygil, Stuart Harrison, Theo Arvanitis

Funding & Partners



A Method for Machine Learning Generation of Realistic Synthetic Datasets for Validating Healthcare Applications

(Arvanitis, White, Harrison, Chaplin, and Despotou, Health Informatics Journal, 2022 in press)

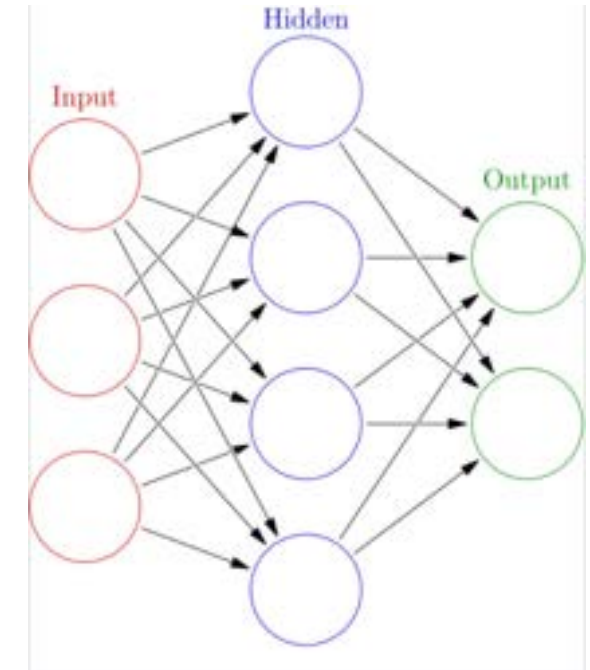
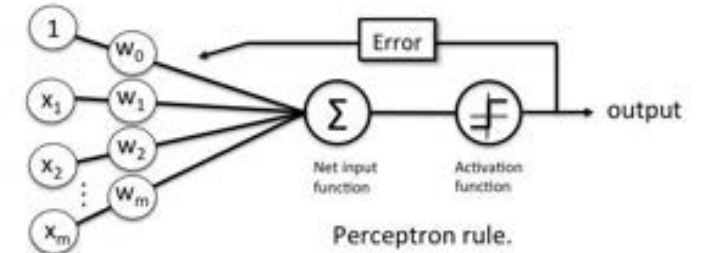


Method – experiment setup:

- Generator produces data
- Discriminator classifies data as fake or real
- Success of the discriminator trains generator to modify data
- Until the discriminator cannot tell
- Both Generator and Discriminator are neural networks: generative adversarial network (GAN) based technology

The experimental setup

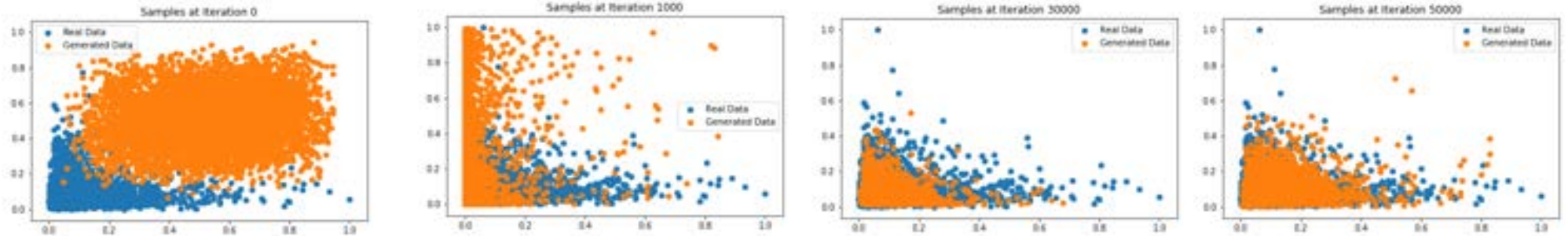
- Neural network configuration consists of a number of elements
 - GAN 2 neural networks, plus feedback from discriminator to generator
 - Activation function: determines when a neuron will 'fire'
 - Epochs: iterations when training the weights and activation function
 - Hidden layers
- Validation of experiment
 - Jaccard similarity measure
 - Jaccard similarity indices are used to compare associations/similarity of two data sets (i.e. ground truth vs synthetic data)
 - Pairwise correlation on the generated dataset and training dataset
 - Correlation matrix



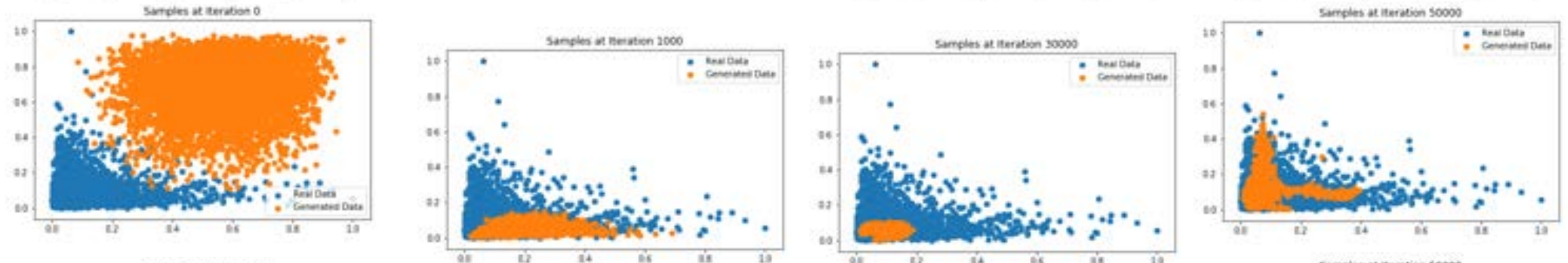
Synthetic Data Generation

MIMIC-III selected numerical data

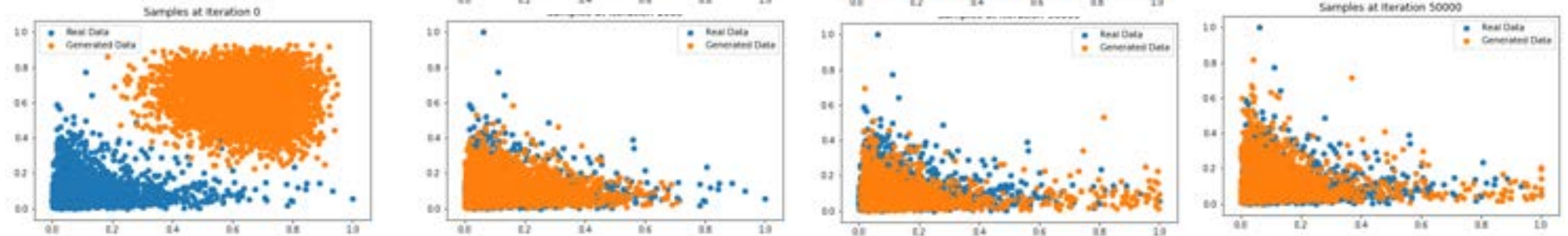
GAN



CGAN



WGAN*

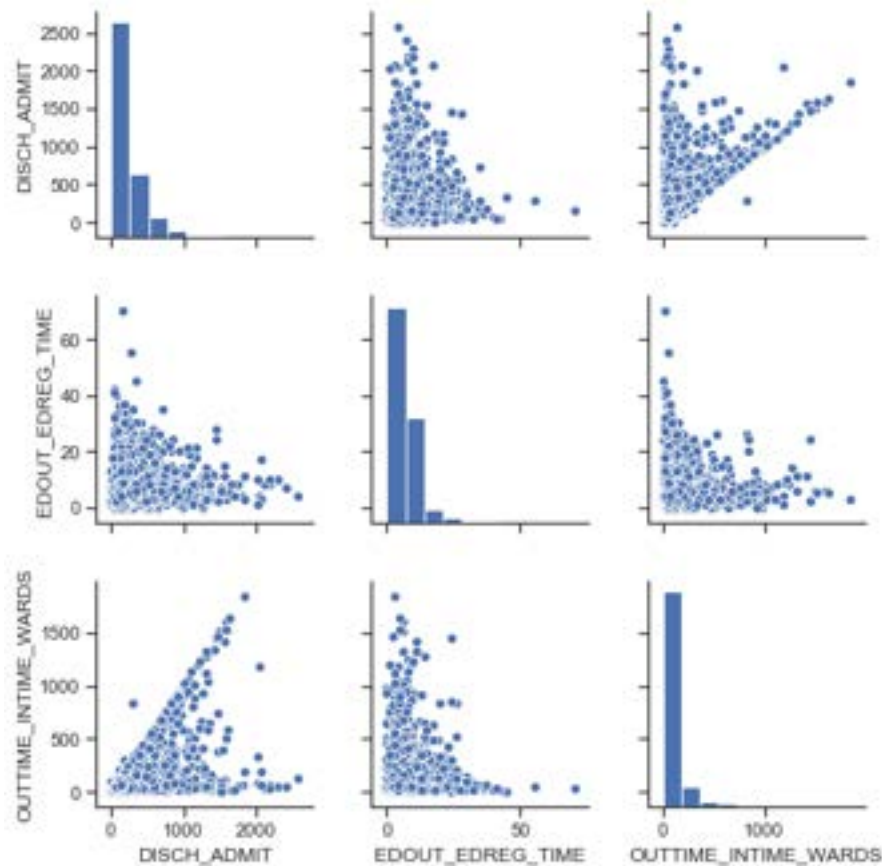


T. N. Arvanitis, S. White, S. Harrison, R. Chaplin, and D. Despotou, A Method for Machine Learning Generation of Realistic Synthetic Datasets for Validating Healthcare Applications, *Health Informatics Journal*, 2022 in press

MIMIC-III data (WGAN)

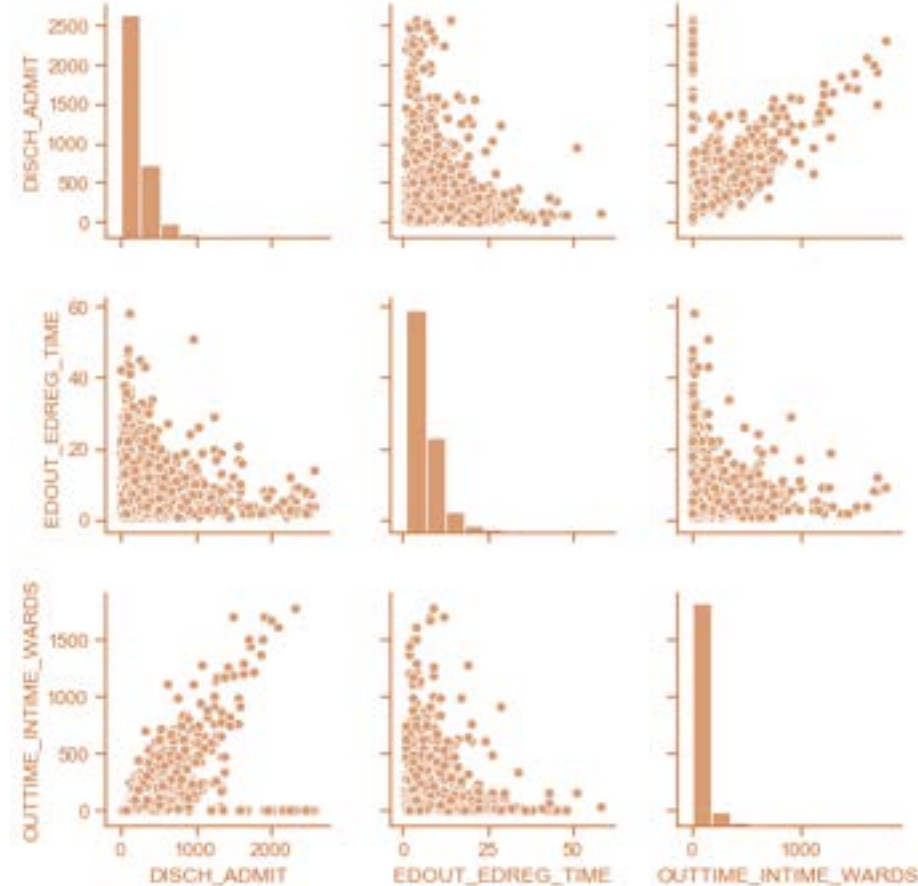
Real data

Pair plot of numerical data



Synthetic data

Pair plot of numerical data



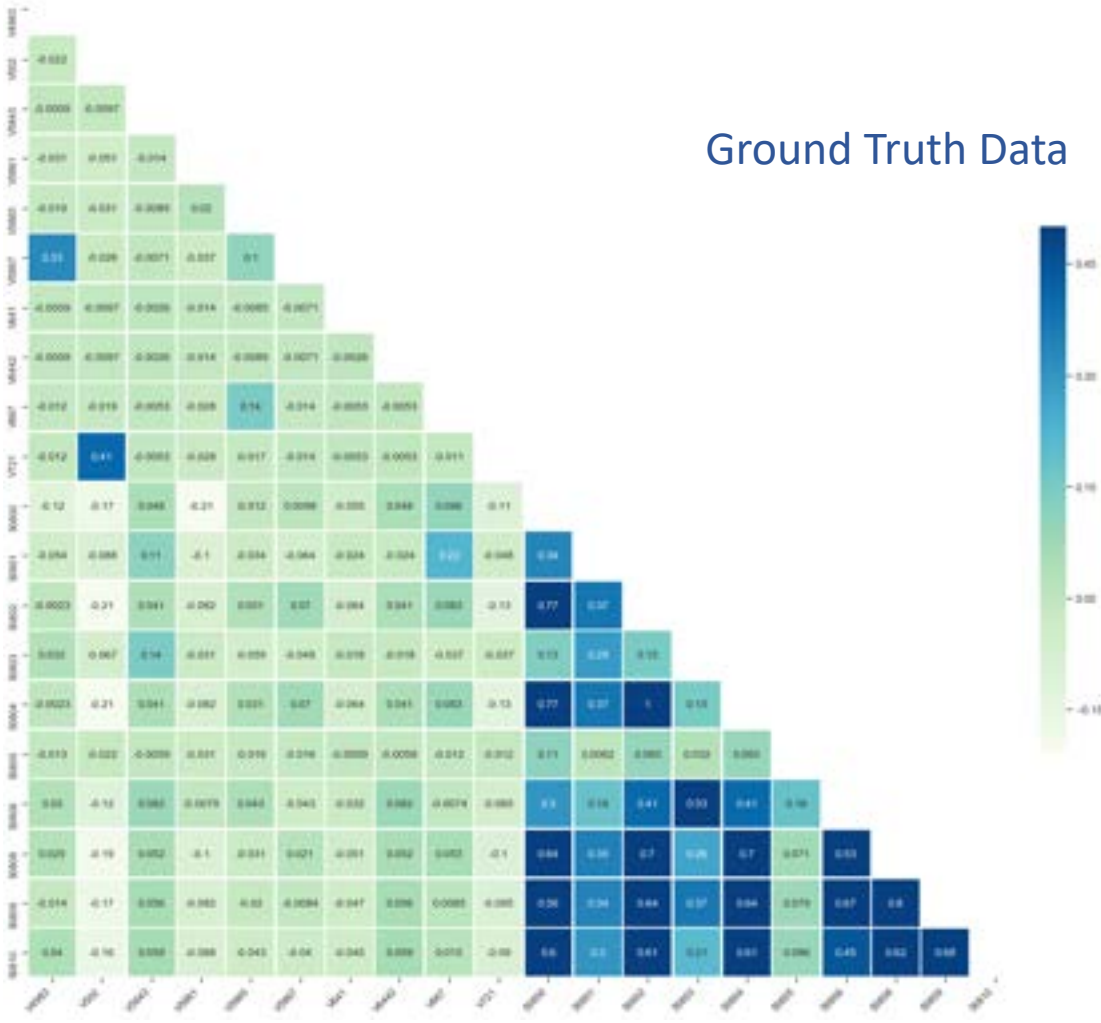
T. N. Arvanitis, S. White, S. Harrison, R. Chaplin, and D. Despotou, A Method for Machine Learning Generation of Realistic Synthetic Datasets for Validating Healthcare Applications, *Health Informatics Journal*, 2022 in press

Synthetic Data Generation

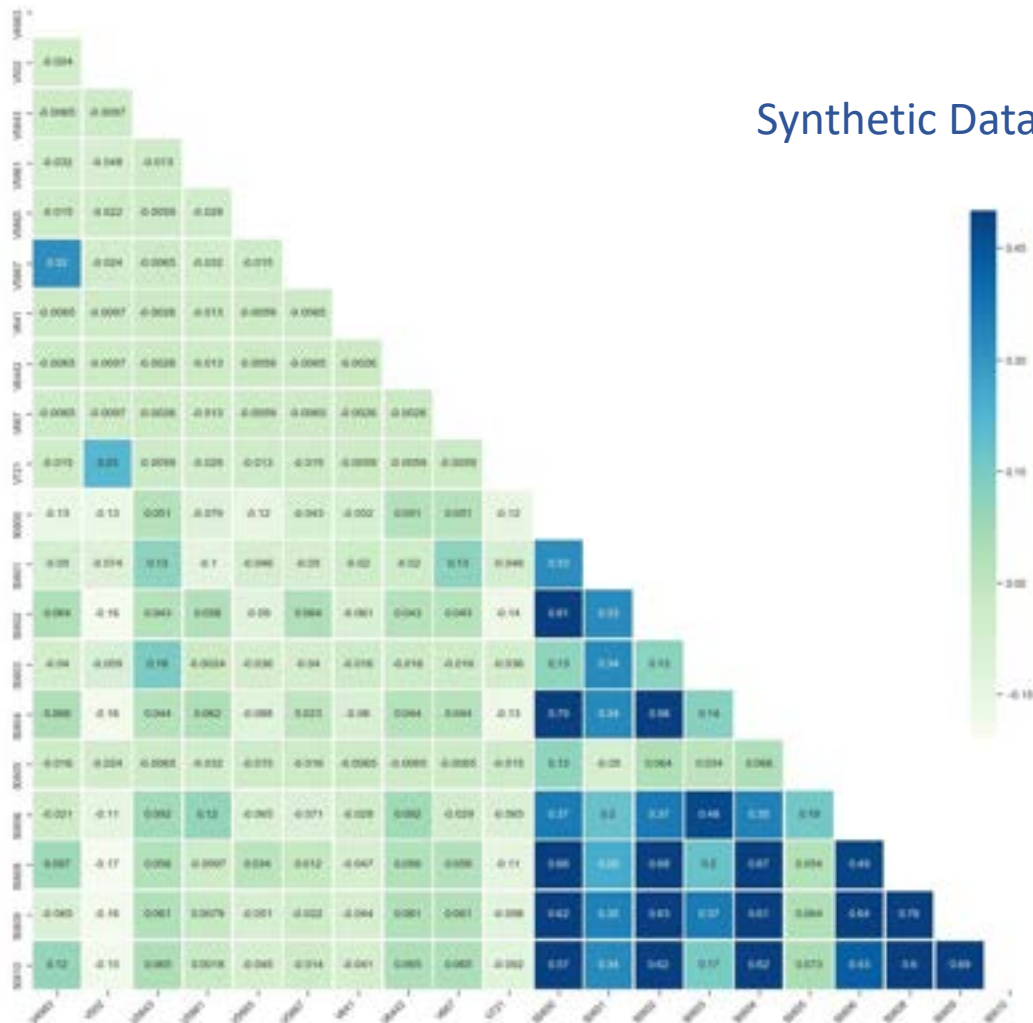
Spearman correlation (10 ICD9 codes, 10 Lab item Codes)

T. N. Arvanitis, S. White, S. Harrison, R. Chaplin, and D. Despotou, A Method for Machine Learning Generation of Realistic Synthetic Datasets for Validating Healthcare Applications, *Health Informatics Journal*, 2022 in press

Ground Truth Data

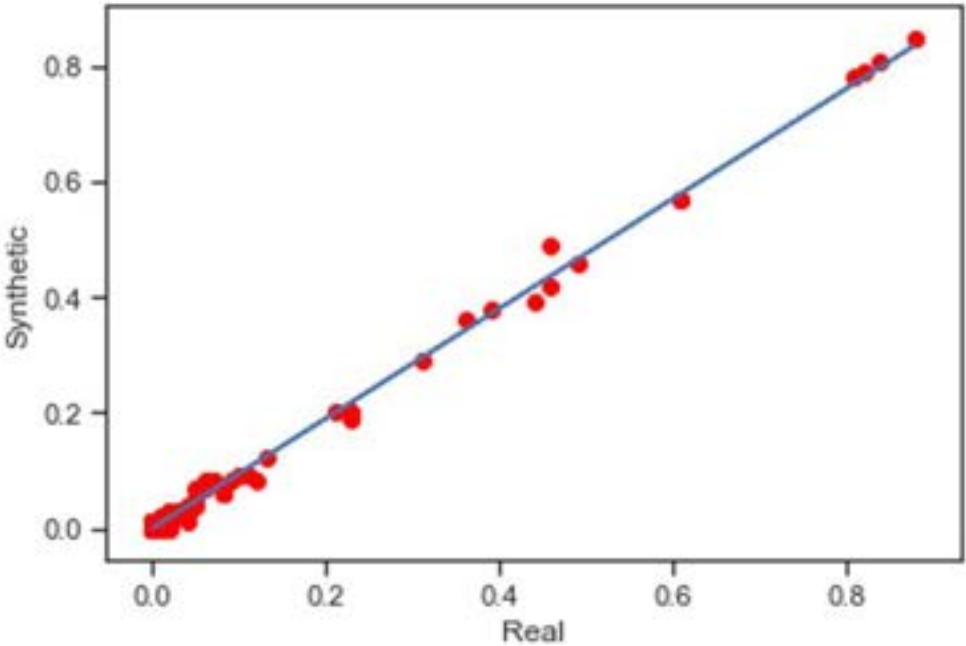
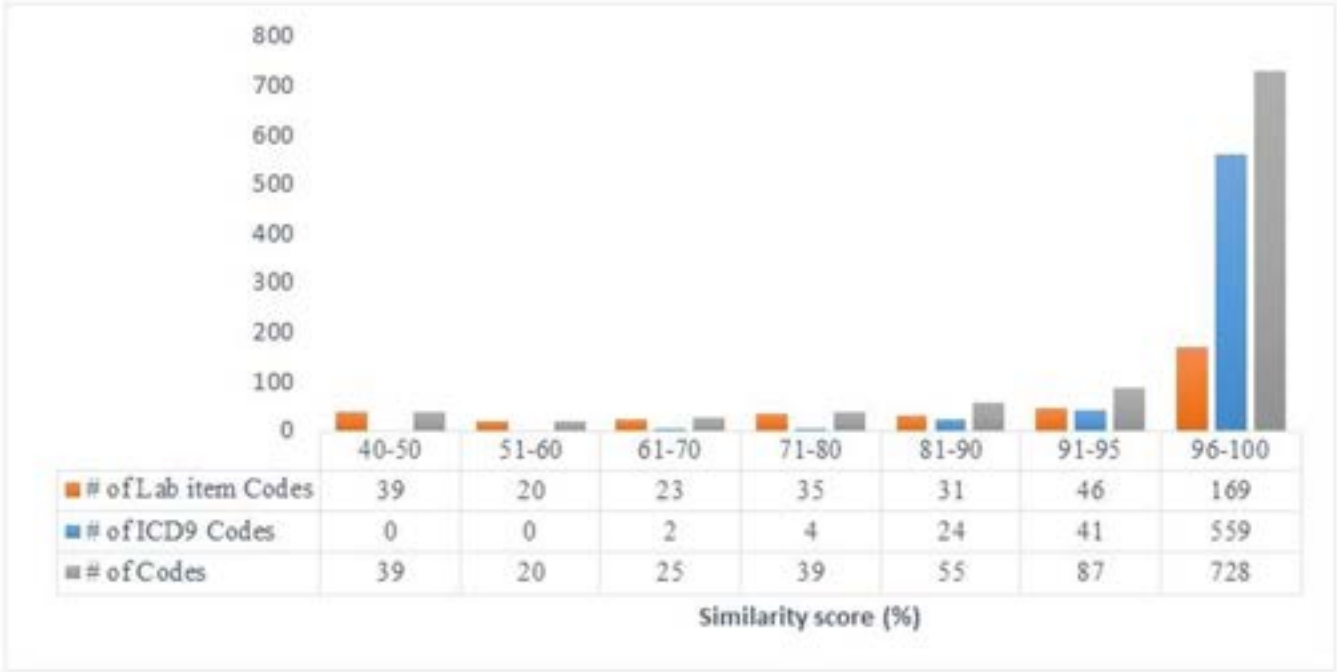


Synthetic Data



Synthetic Data Generation

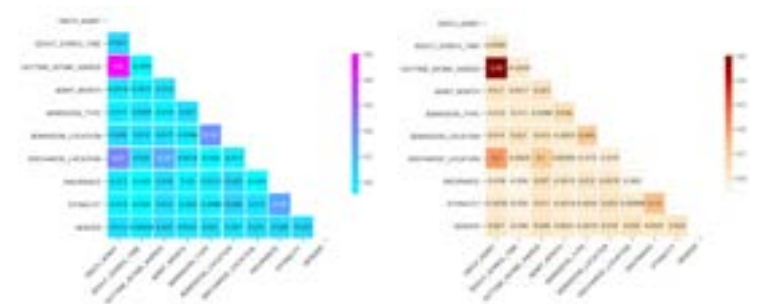
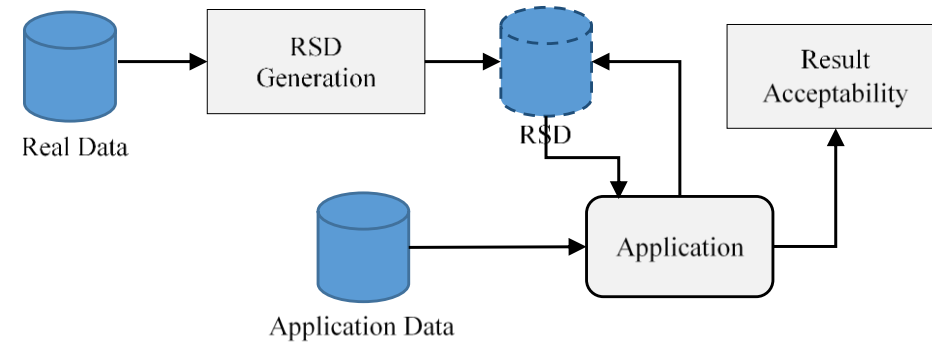
Jaccard similarity of the ICD-9 and lab item codes



T. N. Arvanitis, S. White, S. Harrison, R. Chaplin, and D. Despotou, A Method for Machine Learning Generation of Realistic Synthetic Datasets for Validating Healthcare Applications, *Health Informatics Journal*, 2022 in press

Summary

- **Realistic synthetic health datasets** are an approach recognised as promising, for validation and safety assurance of intelligent healthcare applications.
- Realistic synthetic health data *will overcome barriers of using datasets due to privacy concerns*, enabling development of applications that may increase patient benefit.
- **The GAN based method, successfully generated a realistic synthetic dataset.**
- Statistical tests demonstrated that the two datasets share very similar qualities.
- Some differences between the datasets were identified, particularly with respect to certain lab and ICD-9 codes. This was attributed to low frequency of certain conditions and lab tests. Bigger samples are needed to further explore this aspect.
- Although the datasets share very similar qualities, they are not completely identical. This was a positive finding as it meant that **the GAN did not replicate the real seal dataset values, which would compromise privacy.**



Thank You

T.Arvanitis@warwick.ac.uk

HDRUK
Health Data Research UK



CogStack

Fundamental Infrastructure for 'Unlocking' Electronic Health Record (EHR) data for clinicians, academics and population health analysis.

17/03/2022 | Thomas Searle, King's College London

Supported by the HDRUK Text Analytics Programme



Agenda

- Motivations / Aim
- What is CogStack?
- CogStack: Journey so far:
- CogStack: the road ahead
- Getting involved.

Electronic Health Record Data

- EHR systems hold rich patient data but lack features to ask questions that span across multiple patients in any-where close to real-time
- EHR data can be:
 - Structured (tabular): laboratory test results, basic observations etc.
 - Unstructured (free text): clinical letters, discharge / admission notes, radiology reports etc.

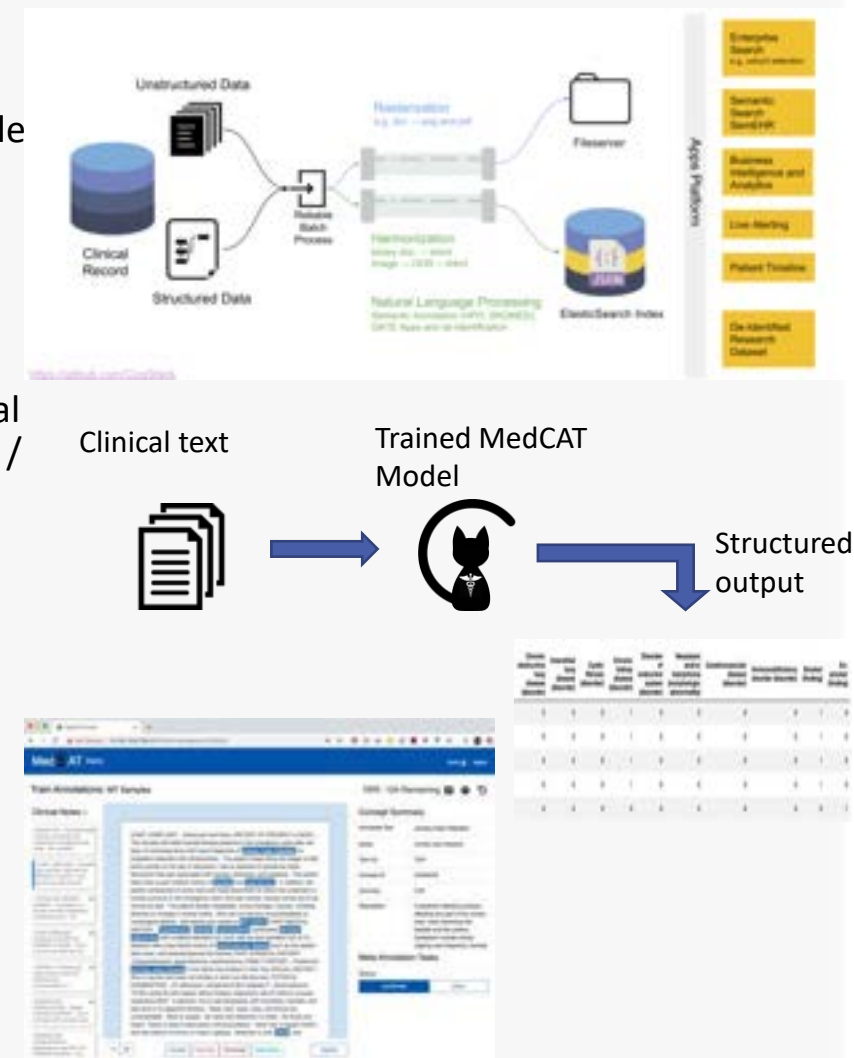
What is CogStack

An ecosystem of loosely coupled technologies that 'unlock' EHR data for:

- Direct patient Care:
 - Real-time alerts
 - Population Health Analysis
- Clinical research
- Clinical Audits
- Clinical Trial Cohort building
- Clinical Coding

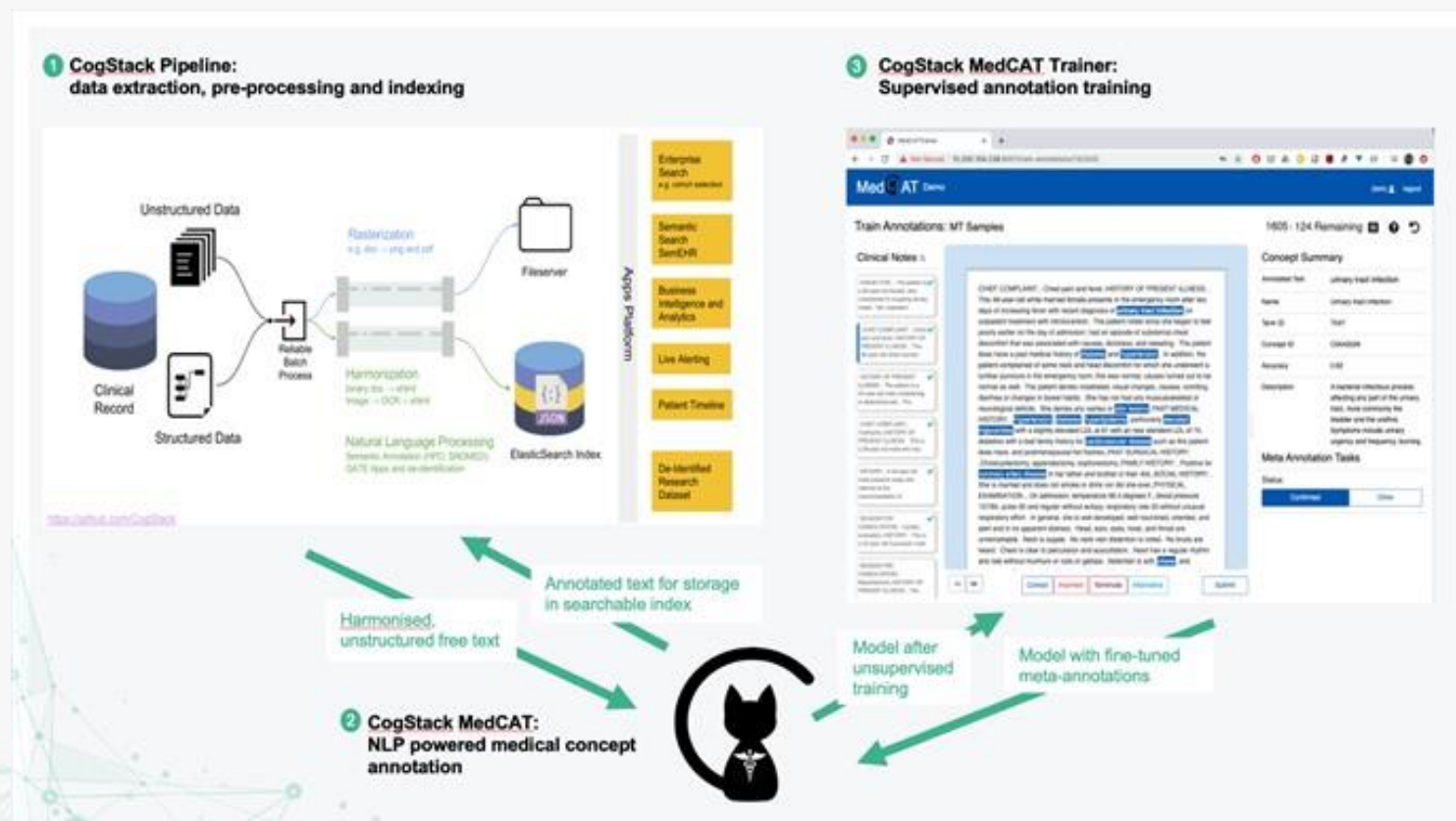
Key Components:

- CogStack-Nifi:
 - Framework for configurable data ingestion pipelines
- MedCAT:
 - A library for training clinical Named Entity Recognition / Linking Models
- MedCATTrainer:
 - An annotation tool for validating, improving and customizing CogStack/MedCAT models



Structuring Electronic Health Records with CogStack / MedCAT

- Structuring EHR data for downstream use
- Novel Natural Language Processing (NLP) models can recognize clinical terms from any terminology: SNOMED-CT, UMSLS, RadLex, ICD, OPCS etc.



Journey So Far

- CogStack has been deployed at:



The road ahead

- NHS AI Lab (formerly NHSx) Stage 3 AI Award:
 - Mature deployments at KCH, UCLH, GSTT, SLaM, UHB
 - Exemplar use case around Clinical Coding
 - Measure potential efficiency / depth of coding gains using CogStack.
- Continue to support deployment specific use cases / improvements.
- Build the community:
 - Open-source / freely available for Trusts / Care Providers to use
 - Grass roots / clinician lead developments
- Prepare fast-follower sites / national rollout.

Thank you / Questions

Supported by the HDRUK Text Analytics Programme



Closing remarks and next steps



Next steps

Meeting follow up



01

- Meeting slides and summary report will be circulated to all attendees
- Please let us know feedback for next time

Events



02

- **04 March:** HDR UK North Digital Care Homes Workshop
- 09 March: CO-CONNECT – COVID - Curated and Open aNalysis aNd rEsearCh plaTform
- **15 March:** Bimonthly Science Webinar
- **16 March:** BHF Data Science Centre Research Showcase
- For more events see the HDR UK website: [Events - HDR UK](#)

Stay in touch



03

- Join the Better Care slack channel (contact alice.turnbull@hdruk.ac.uk)
- Visit the [Better Care webpage](#)
- Visit the [Gateway](#)
- Sign up to the [HDR UK mailing list](#)
- Follow us on LinkedIn and Twitter @HDR_UK

Register now!

BHF Data Science Centre Research Showcase

Wednesday 16th March 2022

Virtual event 10am-4pm



For more details and to register, visit:

<https://www.hdruk.ac.uk/events/bhf-data-science-centre-research-showcase/>

HDRUK
Health Data Research UK



THANK YOU

HDRUK
Health Data Research UK



THANK YOU