

# UCSF

## UC San Francisco Previously Published Works

### Title

A national initiative in data science for health: an evaluation of the UK Farr Institute

### Permalink

<https://escholarship.org/uc/item/93p4m91m>

### Journal

International Journal for Population Data Science, 5(1)

### ISSN

2399-4908

### Authors

Hemingway, H

Lyons, R

Li, Q

et al.

### Publication Date

2020-04-01

### DOI

10.23889/ijpds.v5i1.1128

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

## A national initiative in data science for health: an evaluation of the UK Farr Institute

Hemingway, H<sup>1,2\*</sup>, Lyons, R<sup>3,4</sup>, Li, Q<sup>5,2</sup>, Buchan, I<sup>6</sup>, Ainsworth, J<sup>7</sup>, Pell, J<sup>8</sup>, and Morris, A<sup>9,10</sup>

### Submission History

Submitted:	04/02/2019
Accepted:	26/09/2019
Published:	08/04/2020

<sup>1</sup>HDR UK London<sup>2</sup>UCL Institute of Health Informatics, 222 Euston Road, London NW1 2DA<sup>3</sup>HDRUK Wales/Northern Ireland<sup>4</sup>Swansea University Medical School, Fourth Floor, Data Science Building, Singleton Campus, Swansea, SA2 8PP<sup>5</sup>West China Hospital, Chengdu, China<sup>6</sup>University of Liverpool, Liverpool L69 3BX<sup>7</sup>Division of Informatics, Imaging & Data Sciences, The University of Manchester, Oxford Rd, Manchester M13 9PL<sup>8</sup>Institute of Health and Well-being, University of Glasgow, 1 Lilybank Gardens, Glasgow G12 8RZ<sup>9</sup>University of Edinburgh<sup>10</sup>HDR UK

### Abstract

#### Objective

To evaluate the extent to which the inter-institutional, inter-disciplinary mobilisation of data and skills in the Farr Institute contributed to establishing the emerging field of data science for health in the UK.

#### Design and Outcome measures

We evaluated evidence of six domains characterising a new field of science:

- defining central scientific challenges,
- demonstrating how the central challenges might be solved,
- creating novel interactions among groups of scientists,
- training new types of experts,
- re-organising universities,
- demonstrating impacts in society.

We carried out citation, network and time trend analyses of publications, and a narrative review of infrastructure, methods and tools.

#### Setting

Four UK centres in London, North England, Scotland and Wales (23 university partners), 2013-2018.

#### Results

1. The Farr Institute helped define a central scientific challenge publishing a research corpus, demonstrating insights from electronic health record (EHR) and administrative data at each stage of the translational cycle in 593 papers with at least one Farr Institute author affiliation on PubMed.
2. The Farr Institute offered some demonstrations of how these scientific challenges might be solved: it established the first four ISO27001 certified trusted research environments in the UK, and approved more than 1000 research users, published on 102 unique EHR and administrative data sources, although there was no clear evidence of an increase in novel, sustained record linkages. The Farr Institute established open platforms for the EHR phenotyping algorithms and validations (>70 diseases, CALIBER). Sample sizes showed some evidence of increase but remained less than 10% of the UK population in primary care-hospital care linked studies.
3. The Farr Institute created novel interactions among researchers: the co-author publication network expanded from 944 unique co-authors (based on 67 publications in the first 30 months) to 3839 unique co-authors (545 papers in the final 30 months).
4. Training expanded substantially with 3 new masters courses, training >400 people at masters, short-course and leadership level and 48 PhD students.
5. Universities reorganised with 4/5 Centres established 27 new faculty (tenured) positions, 3 new university institutes.
6. Emerging evidence of impacts included: > 3200 citations for the 10 most cited papers and Farr research informed eight practice-changing clinical guidelines and policies relevant to the health of millions of UK citizens.

#### Conclusion

The Farr Institute played a major role in establishing and growing the field of data science for health in the UK, with some initial evidence of benefits for health and healthcare. The Farr Institute has now expanded into Health Data Research (HDR) UK but key challenges remain including, how to network such activities internationally.

\*Corresponding Author:

Email Address: [h.hemingway@ucl.ac.uk](mailto:h.hemingway@ucl.ac.uk) (H Hemingway)

## What is already known

- National research initiatives in data science for health are underway in several countries seeking to harness insights from electronic health record (EHR) and administrative data at regional and national scale for patient and public benefit.
- One approach to grow this emerging field, adopted by the UK, was to establish a dedicated national research institute, the Farr Institute.
- We do not know how effective such initiatives are. Multi-centre, inter-disciplinary research initiatives are common, but there is a lack of research evaluating such initiatives particularly national research institutes.
- The Farr Institute ran from 2013 until 2018 when its larger-scale successor, Health Data Research (HDR) UK, was established.

## What this study adds

- We provide a framework of six domains relevant for evaluating new inter-institutional, inter-disciplinary initiatives seeking to establish and grow an emerging field of science: defining central scientific challenges, demonstrating how the central challenges might be solved, creating novel interactions among groups of scientists, training new types of experts, re-organising universities, demonstrating impacts in society.
- We show that the Farr Institute created new activities in and across each of the six domains for developing a distinctive research field.
- The Farr Institute demonstrated the ability of multiple UK health research funders and multiple universities to partner in mobilising data, methodology and expertise across disciplines, organisations and information governance domains, resulting in a larger scale of research and improved methodology.
- We have demonstrated globally relevant challenges and opportunities for developing data science for health across institutional and disciplinary barriers, consistent with the need for big investigation not simply big data.
- First, there is a need for a framework by which HDR UK and other national research initiatives might be more rigorously evaluated.

## Introduction

### Rationale for national initiatives in data science for health

Countries across the globe are increasing efforts to harness research insights from EHR and administrative data. A common theme across these initiatives is to access data for research on a bigger scale (number of subjects included in analyses) and with greater detail of clinical and related information, in order to

advance a wide range of research: from disease causation and classification to drug discovery, translational research, clinical trials, evidence-based medicine, clinical practice and public health. While many fields of science have few relevant national borders, data science for health has inherent national and regional dimensions: regional clinical knowledge is required to understand the context and validity of data, and countries differ in the legal, political, economic and public opinion context shaping the research using such data. Countries differ in their approach, for example: nationwide administrative and registry data in Nordic countries [1–4]; province-wide initiatives in Canada [5] and Australia [6], networks of hospitals in the German Medical Informatics Initiative [7] and US PCORnet [8]; and genomics/precision medicine initiatives linked to EHRs such as the US Million Veteran Program [9] and *All of Us* [10].

### UK establishes national research institute

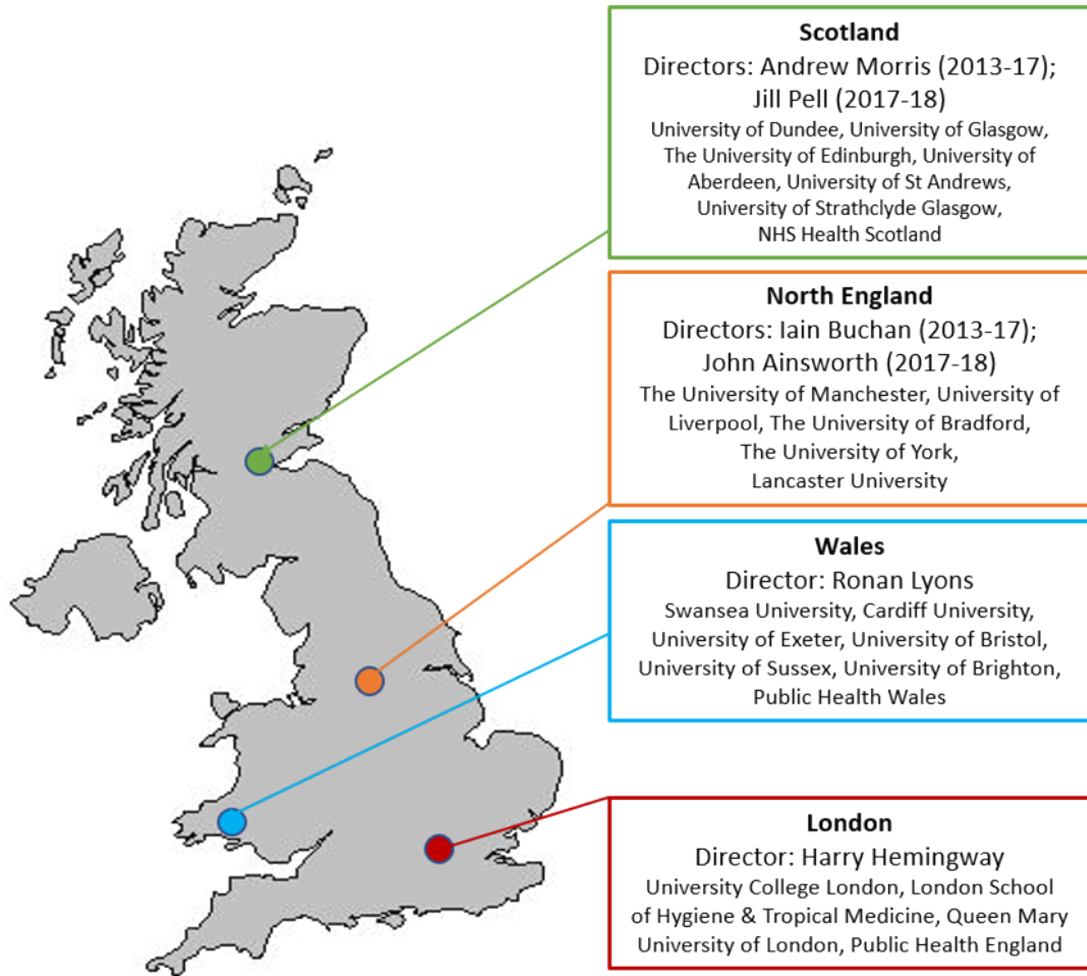
The UK decided a national research institute was necessary to develop and deliver data science for health in order to leverage the NHS and allied data sources. The UK has a population of 65 million, a single payer health system (one NHS), a unique identifier for its citizens' health data, long-standing population-wide EHRs, and 2.2 million citizens in consented cohorts, many with genomic and other detailed research data [11], and an ambition to have 5 million NHS patients with sequenced genomes [12].

The Farr Institute was established in 2013 (supplementary Box 1) to leverage these assets and build capabilities to do research that could only, or best, be done at national scale. The funding for the Farr Institute came from a consortium of ten government and charitable research funders, awarding four academic centres (Scotland, Wales, Northern England and London), involving partnerships across 21 universities in the four centres (Figure 1). In 2018 the Farr Institute made way for the expanded successor institute, HDR UK, with new centres and longer-term core funding (a comparison of the Farr Institute and HDR UK is shown in Table 1).

### Objectives

Empirical evaluations of inter-institutional, inter-disciplinary research initiatives, or indeed national research institutes, have seldom been reported [13, 14]. Evaluations may be conflicted by an interest in persuading funders to continue support. The Farr Institute does not have such a conflict and thus provides an opportunity to learn from its five years of experience. The Farr Institute was a national experiment in how to build the foundations for a new research field of data science for health at scale. In the absence of a well-established framework for evaluating national research institutes, we based our evaluation on the domains of a new field of science, following previous literature [15] and the initial strategy [16] of a research institute in a different field of science, but also launching in 2013 (the Francis Crick Institute).

Figure 1: UK map of four Farr Institute Centres and partners (N=23)



Box: Domains for evaluating new field of science and sources of evidence used

Domain	Source of information
Defining central scientific challenges	Strategy documents Themes, scale, sources, linkages, from 100 most significant publications
Demonstrating how the central challenges might be solved	Survey of Farr Centres: infrastructure, new data made accessible, EHR phenotyping
Novel interactions among groups of scientists	Co-author publication networks, based on all publications with Farr affiliation n=3200
Training new types of experts	Survey of Farr Centres
Re-organising Universities	Survey of Farr Centres
Demonstrating impacts in society	Clinical guidelines and policy documents citing Farr Institute research, changing practice.

## Methods

We evaluated six inter-related domains of a new field of science (see Box), and for each domain sought evidence (publicly accessible and peer reviewed where available) of change during the five years of the Farr Institute (April 2013-March 2018). The evaluation was retrospective and was not planned in 2013.

### Defining central scientific challenges

In order to evaluate how the Farr Institute defined central scientific challenges we analysed published strategy documents and annual reports (available on the Farr Institute website) and invited each Centre to nominate their most scientifically significant publications (25 per centre, 100 total). We (QL and HH) extracted information from these 100 full text publications on attribution to the Farr Institute, different science themes, disease area, scale, number of data sources, new record linkages, and evidence of science impact from citation tracking (Google Scholar, accessed September 2019).

To be eligible, each publication reported the use of one or more source of EHR or administrative data or methods directly relevant to data science. Consented studies without use of such EHR or administrative data were not eligible. We classified research themes (following strategy documents) as: citizen driven health, discovery science, quality and outcomes, trials and public health. We defined attribution to the Farr Institute as at least one author who: listed Farr Institute as an author affiliation in PubMed, acknowledged funding for the Farr Institute, or was in receipt of funding from the Farr Institute. Data sources were classified as primary care, hospital discharge data (e.g., Hospital Episode Statistics for England, Patient Episode Database for Wales, Scottish Morbidity Record), detailed hospital data, disease and procedure registries, mortality, other health, and socio-economic and other non-health data. We extracted for each publication the number of people providing the denominator population (sample size) and classified the population as healthy (general population sample) or based on specific disease or procedure. We carried out a structured survey with each Centre completing information on how the central challenges might be solved in three areas: e-infrastructure and platforms for accessing and sharing data, new data sources made available for researchers, and EHR phenotyping methods for structured and unstructured data.

For novel interactions among groups of scientists we analysed peer reviewed publications. In addition to the in-depth analysis of the self nominated 100 most significant publications, we also sought to identify all publications with at least one author listing a Farr affiliation. We searched PubMed from inception to 18 September 2018, using the strategy 'Farr Institute[AD]'. We visualised the change over time in scientist network behaviour in the Farr Institute, with co-author publications networks based on all publications returned from PubMed, using Cytoscape cross-institution activity by year of publication. We analysed co-author publication network and inter-disciplinarity by the halfway time point, comparing the first vs the final 30 months of the Farr Institute.

For training new types of experts and Re-organising Universities, we carried out a structured survey across centres.

For impact in society, we identified clinical guidelines and policy documents citing Farr Institute publications through an-

nual reports to funders, automated software used by funders to capture a range of outputs and impacts (Researchfish) and by contacting investigators.

## Results

### Defining central scientific challenges

The Farr Institute published a research corpus around a central scientific challenge, demonstrating insights from EHRs and administrative data. Figure 2 (top panel) illustrates the challenges of the Farr Institute and (middle panel) demonstrates the corresponding research publication corpus. The proportion (%) of publications at different stages of the translational cycle was: methods (17%), patient involvement (3%), discovery science (8%), health services research (24%), clinical trials (5%) and public health (42%). The Farr Institute published highly-cited papers using EHR at each stage in the translational cycle (the top 10 are shown in Table 2), with a total of 3200 citations. Four of these highly-cited papers illustrate the higher resolution of using linked EHR [17–20]. These different research themes were applied across different clinical areas, including cardiometabolic, maternity and child health, mental health, cancer, renal and respiratory.

There was some evidence of a modest increase in the scale (number of people analysed in the denominator population) of research over time in these publications (Figure 3), based on linked primary-secondary care data in adults. But by 2018 this represented only 6.15% of the UK population [21]. There was just one paper that used the whole of England's hospitalisation data: Freemantle and colleagues [22] analysing weekend mortality effects using 14.8 million admissions, and several using all England's deaths data.

### Demonstrating how the central challenges might be solved: access to research-ready data

In 2013 there were no independently accredited Trusted Research Environments (TRE) for NHS data: by 2018 there were four (one in each centre) ISO 27001:2013 certified data safe havens (Table 3a). The TREs provide secure remote access, a safe environment for the analysis of sensitive patient identifiable data, a pre-requisite for receiving unconsented, individual level health data for research use. We found evidence that Farr activity enabled other scientific fields: with over 1000 approved users on these four data safe havens working on over 300 research projects (the majority being external, having no Farr Institute funding). The Farr Institute enabled the research use of diverse anonymised patient records, linked across primary care and secondary care, including NHS imaging data, blood laboratory values and reimbursed prescriptions (Table 3b). There was a cumulative total of 102 unique data sources reported in these publications (Figure 4a), with 13% from primary care, 17% limited coded hospital data, 8% detailed hospital data, 19% registries of disease and procedures, 26% socio-economic and environment, 6% death data and 12% other health data. The setting and names of each unique data source reported in these publications are shown in Supplementary Table 1. We



Figure 2: Central scientific challenges of the Farr Institute



found that the median number of record sources per publication showed no evidence of increasing, and if anything decreased over the 5 years (Figure 4b).

### Demonstrating how the central challenges might be solved: phenotyping

In 2013 there were no openly accessible portals for defining diseases and health-related conditions using EHR data (EHR ‘phenotyping’). The Farr Institute supported several initiatives in disease phenotyping (Table 3c): these included CALIBER, an open platform [23] of re-usable EHR phenotypes (code lists + logic + validations) for over 70 diseases which have been re-used in more than 50 publications with more than 80 ongoing projects [24]. In addition, there were several publications of EHR phenotypes in Wales and Scotland [25] and a clinical code repository [26].

Methods of surfacing the entire structured and unstructured data in a hospital have now been demonstrated in three hospitals with CogStack and SemEHR [27].

### Creating Novel interactions among groups of scientists

Based on author affiliation, the search [Farr Institute[AD]] on PubMed returned 594 unique publications (from inception to 18 September 2018). Figure 5 shows that there was a large expansion of co-author publication networks comparing the first 30 months (67 publications with 944 unique co-authors) and the final 30 months (545 papers and 3839 unique co-

authors). Supplementary Figure 4 shows that overall across the 100 publications, 28% included Farr Institute as both author affiliation and funder acknowledgement, 14% as author affiliation only, 11% as funder acknowledgement only and 42% as Farr-funded investigator only, as confirmed by the centres. There was some evidence that over time both author affiliation and funder acknowledgement increased. Based on the departmental affiliation of co-authors there was some evidence of greater inter-disciplinarity in the last 30 months of the Farr Institute compared to the initial 30 months (Supplementary Figure 5). We identified 17% of publications involving universities from across two or more Farr Centres.

We explored the number of publications among the top 100 papers with an international (non-UK) author affiliation and found the total number of publications as follows: 17 publications non-Nordic Europe, 11 US, 10 Nordic, 8 Canada, 4 Australia and 2 China.

Among the 100 papers identified by the Farr Institute, 87 were found in the Scopus database, which provides structured institution data used to generate Supplementary Figure 6. The median (IQR) number of universities per Farr Institute publication was 3 (IQR: 2-4) in 2013 and 6 (IQR: 4-8) in 2018 (Supplementary Figure 6). In April 2017 the Farr Institute with the European Federation of Medical Informatics attracted over 850 delegates from around the world to Informatics for Health [28] – the UK’s largest health and biomedical informatics research gathering to date.

Figure 3: Sample size as a percentage of total UK population reported in research publications

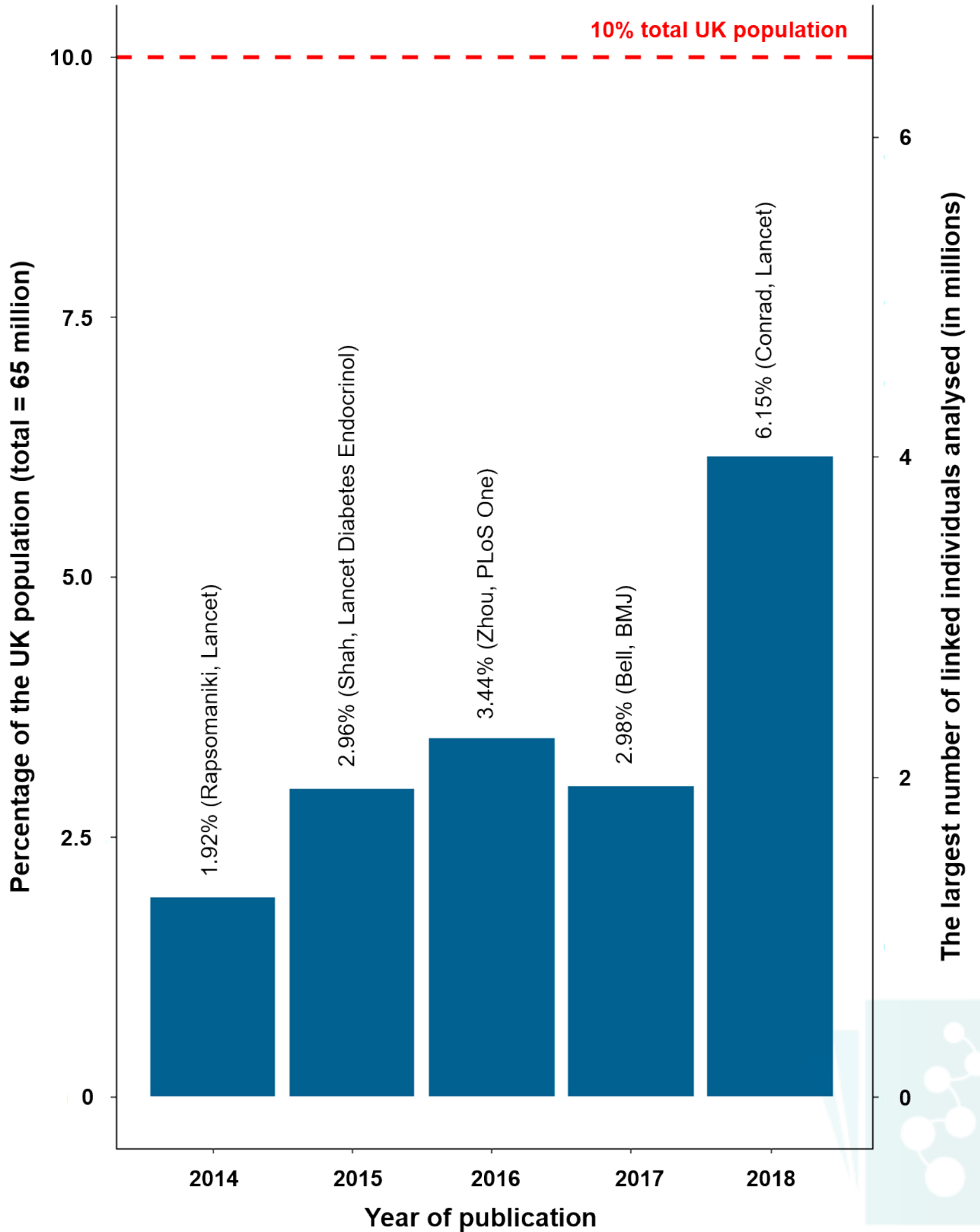


Figure 4a: Cumulative total of unique electronic health record data sources reported in Farr Institute publications

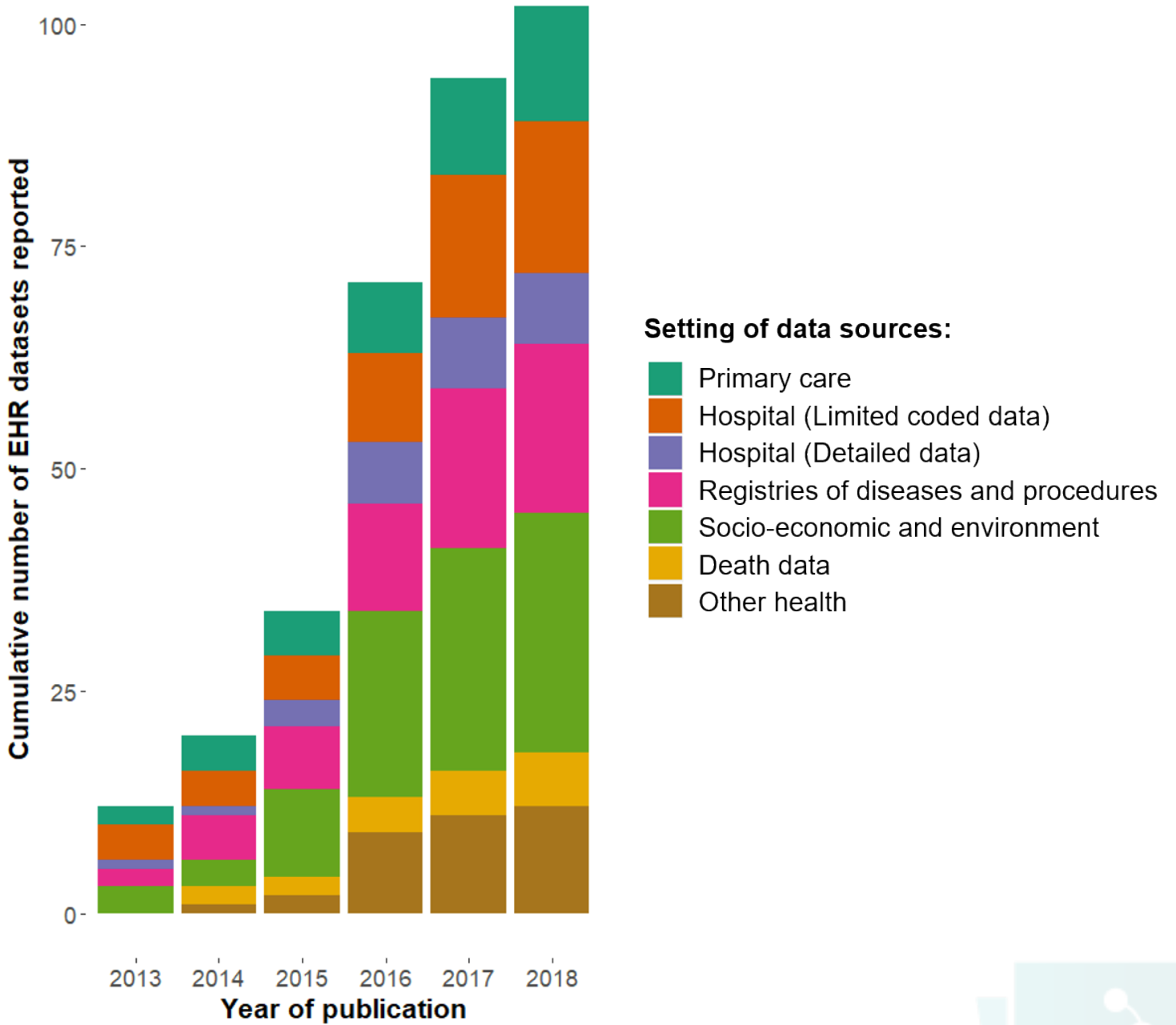




Figure 4b: Number of electronic health record data sources reported in each publication (from Farr top 100)

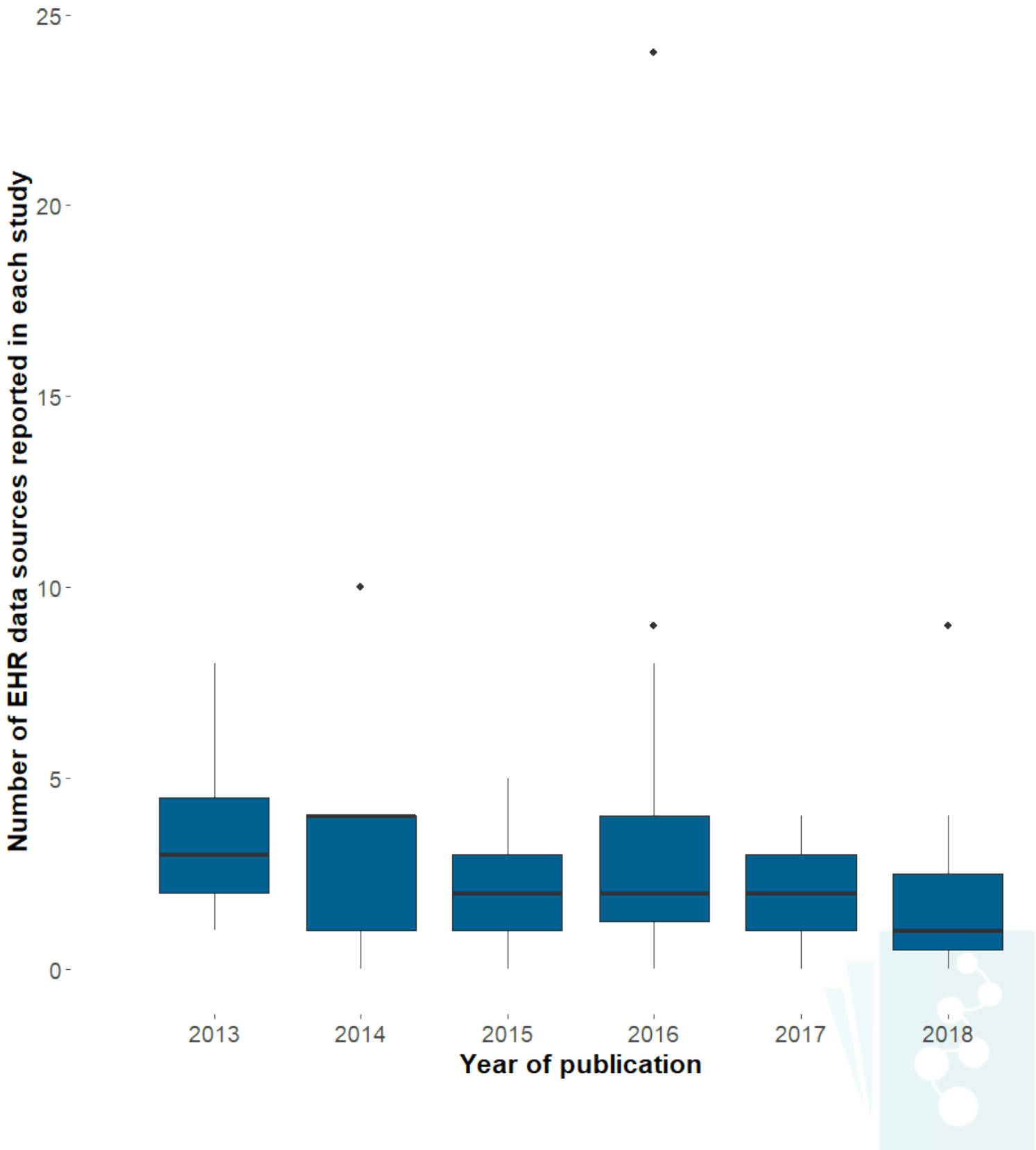
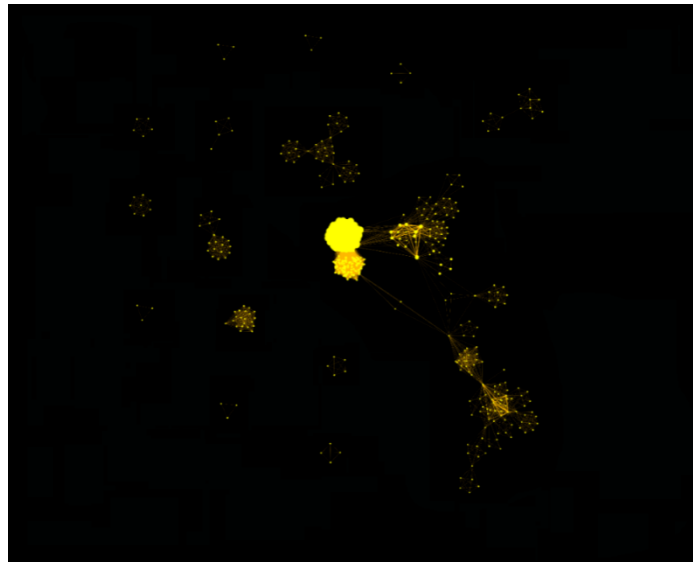


Figure 5: Change in co-author publication networks between the first and final 30-month periods (each dot represents a unique author: lines connect authors publishing together)

a) First 30 months



b) final 30 months



## Training new types of experts

The Farr Institute Centres created eight new training programmes at masters, short-course and leadership level across the four Farr Centres (Table 4). In total 2102 people were trained in informatics and directly relevant courses across the 4 Centres. As well as conventional programmes - there were 48 PhD students directly funded by the Farr Institute – the Farr Institute also established a competitively awarded Farr Future Leaders residential programme with two cohorts of 21 mid-career researchers with facilitation of national research collaborations in conjunction with leadership training.

## Re-organising Universities

Universities associated with the Farr Institute established new institutes and structures (cost centres) for health informatics and data science. At UCL (Institute of Health Informatics, 2014), University of Manchester (Division of Informatics, Imaging and Data Sciences, Health eResearch Centre), Edinburgh (Usher Institute of Population Sciences and Informatics), Swansea University (Patient and Population Health Informatics). These universities established 24 Faculty (tenured) posts as enduring leverage of the Farr award.

## Demonstrating impacts in society

Farr Institute research at different stages of the translational cycle was cited in and informed practice-changing clinical guidelines and policies Table 5. These examples included research in public health (WHO guideline recommendations on video observed therapy for TB), clinical risk prediction (European Society of Cardiology), drug regulatory approvals (NICE potentially affecting the type or duration of drug treatment), and implementation of genomic medicine. The change in practice recommendations potentially affects more than a million UK citizens.

More widely Farr informatics research informed changes in government strategy from a centralised to a decentralised approach to integrating place-based health and administrative data for multiple analytic purposes [29]. This work also generated a £20m pilot of problem-based data integration, pulling through data by addressing care pathway blockages and research questions of importance to the local community in regions of 3-5m population [30]. This became the blueprint for England's Local Health and Care Record Exemplars [31].

## Discussion

Clinicians, patients and policy makers have growing expectations of the use of data to provide research insights with the potential to improve health and care outcomes [32]. New scientific fields tend to have high-priority defining characteristics; we provide evidence in six recognized domains suggesting the Farr Institute played a major role in establishing and growing the field of data science for health in the UK. The experience of the Farr Institute has informed the design of Health Data Research UK (HDR UK), and this evaluation is relevant to the inter-institution, inter-disciplinary challenges of scaling up health science around big data in many parts of the world.

## Evolution of UK national research institute in policy context

A substantial achievement of the Farr Institute and its funders was the founding of its larger successor HDR UK. The key differences and similarities of the two organisations are shown in Table 1. This expansion from Farr Institute to HDR UK was recommended by the Farr Institute International Advisory Board in 2015, and much of the evolution of the Farr (see timeline in Supplementary Fig 3) was in preparation for this change. Lessons learned from the Farr Institute informed the development of HDR UK's science strategy [33] (with 6 themes in human phenome project, AI, multiomics and multi-disease aetiology, clinical trials, digital health insights, and public health) and delivery, establishing a new single legal entity, single Director and Board, working via consortium agreements. The expanded research organisation partnerships in HDR UK now includes the majority of UK research organisations with significant expertise in data science for health. During 2013-18 the UK underwent a set of profound political, societal and policy changes relevant to data science for health: implemented the most significant reforms of the NHS of a generation in 2013 (establishing NHS Digital, Public Health England), enduring multiple crises of public trust over the use of data (care.data, Royal Free NHS Foundation Trust and Google DeepMind), the most significant changes in data protection legislation in a generation (as a result of the EU General Data Protection Regulation), two General Elections, and the historic referendum to leave the EU (see timeline in Supplementary Fig 4).

## International context of national research initiatives in data science for health

The Farr Institute and HDR UK have sought to learn from the differing approaches among countries and jurisdictions to advancing data science for health. For example, in order to facilitate learning across national boundaries, the Farr Institute jointly funded exchange fellowships with the Institute of Clinical Evaluative Sciences in Ontario, Canada. Currently, as far as the authors are aware, other countries have not established a national research institute dedicated to data science for health directly analogous to the Farr Institute or HDR UK. There is a need for strategic collaboration across national borders in data science for health [34], with proposals for working across national borders in Nordic countries [4].

The UK, informed by the ability of smaller countries such as Denmark [35] to deliver nationwide research built on an array of linkable record resources, as established the national institute in order to meet the challenging goal of nationwide research in a country of 65 million people. Some of the challenges facing the Farr Institute, and now HDR UK, are common to any research initiative based on catalysing inter-institutional and inter-disciplinary collaboration. Previous policy reports have recommended the need for intra-national methodological developments in data science for health as an important basis for international collaboration [36].

## Defining central scientific challenges: biomedical

The central scientific challenge set out by the 2011 funding competition (supplementary MRC 2011 Funding call ) concerned record linkages. After the Farr Institute was established, the international advisory board, funders and directors recognised the need to more clearly identify the key research challenges which could only or best be addressed with centres coming together as a single national research institute. This led, in 2015, to the development of the science themes of discovery science, precision medicine, trials, learning health systems and public health outlined in Figure 2.

## Defining central scientific challenges: scale

Providing a ‘more powerful telescope’ by enabling EHR and administrative data at greater scale (larger sample sizes) is part of the central scientific challenge. Although nationwide primary care data exist in the UK with the 65 million population held by four GP system suppliers, they are not brought together in a single dataset for research purposes. The Farr Institute was publishing on 6% samples at 2018. The largest collections are held by Q-Research, Clinical Practice Research Datalink (CPRD) and System One, and there are more extensive collections of linked GP and other care data in the regions and devolved administrations, such as the Secure Anonymised Information Linkage (SAIL) system in Wales and the Connected Health Cities’ Trustworthy Research Environments in North England’s four regions. The levers for increasing scale do not sit in research institutes, but with data controllers and the legal and political environments. The Farr Institute paved the way for federation of research data queries and distributed analytics across regional data aggregations.

## Defining central scientific challenges: across the translational cycle

Most biomedical research disciplines are focused on a particular phase of the translational cycle: the Farr Institute demonstrated that a distinctive contribution of data science for health is that EHR and other sources of data ‘in the wild’ can link investigators across all phases of the translational research cycle. The Farr Institute made a start in the UK: the ambition, which HDR UK has taken on, is to constructively disrupt current models of evidence-based medicine, clinical practice and translational research, including the way that research is organised and funded.

## Defining central scientific challenges: record linkages

The original 2011 funding competition (supplementary MRC funding call document) sought centres to “undertake and promote innovative linkage and analysis of large health related data sets”. The top 100 Farr Institute publications reported 102 unique data sets, but record linkages were not reported consistently, nor in sufficient detail to know which sources had been linked, and whether any linkage was new. In Wales (SAIL), Scotland (eIDRIS) and some English regions (Connected Health Cities) there are data linkage and trustworthy

research environments that have fuelled numerous research outputs. For example, in Wales primary care data (including text) are linked to hospital admissions data, dispensed prescribing, blood laboratory values and a wide range of socio-economic data. This breadth and depth of linkages, and their sustained accessibility by researchers, have not emerged across larger populations such as England. In England the opportunities for developing a growing, sustainable environment for record linkages were severely curtailed by care.data and have only recovered in the regional devolved approaches such as the NHS England Local Health and Care Record Exemplars. In annotating Farr Institute research publications, we found variable clarity on reporting of record linkage and were not able to easily identify how many linkages had been reported which were new and which might be readily accessible to future researchers.

## Defining central scientific challenges: data quality

Improvements in the quality of health record data, an internationally recognised challenge [35], was not an aim of the original funding competition, nor did it become a co-ordinated national focus in the Farr Institute, in part because the key influences lie in the NHS. Nonetheless the Farr Institute did deliver an open portal, CALIBER, for EHR phenotyping algorithms, along with evaluations of data quality and validations.

## Demonstrating how the central challenges might be solved

The Farr Institute transformed the UK’s ability to bring non-consented individual-level health data into trustworthy environments and make them available for other researchers, based on specifically approved projects. Nonetheless, there remain many different data governance environments and processes for data access for research, with much room for harmonisation and streamlining. We demonstrate here how the Farr Institute published on over 100 EHR and administrative data sources; in some situations these were the first research use of these data. Despite the undoubted progress reported here, the EHR data sources reported represent a tiny proportion of available data.

## How the challenge may be addressed: phenotyping

The Farr Institute demonstrated approaches to the major challenge of defining disease and health related traits with diverse EHR and other data. The development of the open CALIBER portal for EHR phenotype algorithms and their validations provides a basis on which to develop a national online facility to integrate data, methods and investigators for EHR phenotyping. Recent work from the Farr Institute demonstrated multiple code set validations of the 308 most common diseases and conditions, providing a ‘chronological map’ of human health from birth to death [37,38]. A key challenge for the future is to develop approaches which provide a degree of consistency and replicability across jurisdictions and national borders.

## Creating novel interactions among groups of scientists

We visualised an 'explosion' of co-author networks. This reflects the willingness of investigators to self-identify with the Farr Institute, as there was no monitoring of this practice at central or national level, as well as extensive collaborations between those with and without Farr Institute funding. The top ten most cited Farr publications have been cited in total >3200 times. The capital investment allowed the co-ordinated establishment, for the first time in the UK, of physical homes (buildings) for the emerging field: 6421m<sup>2</sup> of dedicated, branded space allowing research disciplines, technical specialists, NHS IT professionals and industry partners to 'breathe the same air'. Beyond 2018, the capital investment in the Farr Institute has had an enduring effect, with most of the physical estate now being dedicated to HDR UK activities. The Farr Institute co-sponsored substantial expansion of scientific conferences: >850 delegates from 13 countries at Manchester 2017 meeting (>3 times larger than 2013 conference). Two new journals were established (*International Journal of Population Data Science* and *Journal of Learning Health Systems*) with Farr investigators Kerina Jones and Iain Buchan as editors.

There remain major challenges in bringing scientists and technical specialists to work together across different institutions and disciplines. Indeed the influential Research Excellence Framework, provides financial incentives to universities to compete, rather than collaborate [39].

## Impact on Innovation and Industry Partnerships

Universities within the Farr Institute Centres collaborated with industrial partners on research (e.g. Astra Zeneca, see Table 5), and training (e.g. masters courses with an option for students to complete their dissertation as interns in industry). However with 21 university partners the Farr Institute was not able to sign strategic industry partnerships.

## Training new types of experts

At masters, short-course and leadership levels, the Farr Institute had a substantial effect on teaching and training. The Farr Institute evolved its training opportunities in response to the growing need for data science as a discipline and in leadership, neither of which were explicitly envisaged in the 2011 original funding call (supplementary MRC 2011 Funding call).

## Demonstrating impacts in society: on health and healthcare

The Farr Institute carried out research underpinning policies and recommendations to change clinical and public health practice, and shaping government policy in health data management and digital health innovation. We provide here examples of specific research findings and their relation to changes in policy and recommendations. However, the Farr Institute had no central mechanism of identifying such influence (Table 5 is likely incomplete), nor of prospectively following research through policy recommendations to measure changes in health. In some cases Farr research may have impacts in later

years; HDR UK might usefully establish a more systematic approach.

## Demonstrating impacts in society: public engagement and trust

The Farr Institute directly funded the involvement and engagement of the public and patients in research using patient data. In 2013 there were no national campaigns involving patients and the public in research on patient data. By 2018, the Farr Institute had delivered influential national campaigns in public engagement with the '100 ways' case studies, explaining to patients and public examples of the benefits of Farr Institute research (13,000 followers and subscribers), and the #datasaveslives campaign, which has generated more than 50,000 retweets. In addition, the Farr Institute contributed to the Understanding Patient Data Wellcome Trust initiative [40]. The Farr Institute established public panels, a network of over 50 regular public contributors, and using a range of deliberative methods engaged members of the public in dialogue around the ways that data are used in health research; consulting them on research and governance practices and involving them in agenda-setting and decision-making processes.

## Limitations and Challenges

We identify five important limitations to this evaluation and briefly discuss how these limitations might be addressed for others evaluating inter-institutional, inter-disciplinary research initiatives. First, the challenge of attribution: would the changes in each domain which we catalogue here have happened anyway, in the absence of the Farr Institute? Because the funding of the Farr was intended to catalyse and leverage a range of activities, rather than to wholly fund specific activities, it is seldom possible to prove attribution robustly. This challenge of attribution is illustrated with research publications. In nearly all research publications analysed in this evaluation, the Farr Institute was one of several funders, and co-authors with Farr funding are usually among a larger group of co-authors without Farr funding. The Farr Institute did not have an intra-mural programme of research fully funded by the Farr awards, which were initially made to build four centres. Unsurprisingly therefore, the major peer-reviewed scientific outputs of the Farr were not published by all four centres together. This illustrates the challenge of transparent and publicly accessible attribution to a national institute.

Second, the Farr Institute did not design a prospective evaluation at the outset and thus the evidence presented here has inherent limitations of a post hoc evaluation. A key lesson learned was the importance of establishing at the outset national data systems allowing research about research ideally in near real time. These key data elements include items that currently do have unique identifiers: Farr investigators (identified by ORCID iDs), their publications (identified by PMID), but also the Farr Institute could have done more to advance central cataloguing (with unique identifiers) of datasets, record linkages, projects, disciplines, employing organisation and department, partner organisations (e.g. NHS). Indeed, much of the evidence reported in this evaluation could not have been identified by others, and some of it remains not open to independent scrutiny.



Third, evaluations require comparisons and ours was limited to the first and final 30 months (a 'before and after' design). We believe it would be informative in future to compare with other inter-institutional, inter-disciplinary national research institutes (including for example in the UK, the Crick, Turing, Sanger, Dementia Research Institute).

Fourth, our evaluation was neither external to nor independent of, the Farr Institute. However, the Farr Institute did receive regular external feedback, from both the International Advisory Board and the funder consortium. Since the Farr Institute ended in 2018, there were no potential conflicts of trying to seek further funding.

Fifth, we recognize that there is a role for qualitative evidence, for example through interviews with stakeholders as well as critics, which might add insights and highlight further critical challenges on many of the domains which we sought to evaluate, including the nature, and extent, of any changes attributable to the Farr Institute.

## Conclusion

In the UK, the Farr Institute played a significant role in beginning to grow the field of data science for health. In 2013 there was little UK-wide co-ordination or visibility of research capabilities, including methods development or training in data science for health, and by 2018 this had been transformed. The importance of a national research institute in this field is evidenced by the UK's expanded commitment to HDR UK.

## Acknowledgements

The funders of the Farr Institute were: Medical Research Council (co-ordinating), Arthritis Research UK, British Heart Foundation, Cancer Research UK, Chief Scientist Office, Economic and Social Research Council, Engineering and Physical Sciences Research Council, National Institute for Health Research, National Institute for Social Care and Health Research, and Wellcome Trust.

## Farr Institute

The work of the Farr Institute was carried out by a network of about 430 investigators, students and professional support staff: the majority of whom received either no, or limited direct funding support from the Farr Institute award for their salaries.

## Conflict of interest statement

The authors report no conflicts of interest.

## Ethics statement

Ethical committee approval was not required for this evaluation.

## References

1. Swedish National Data Service "Register-based research" [Online]. Available: <https://snd.gu.se/en/data-management/register-based-research>. [Accessed: 21-Nov-2018].
2. L. Leitsalu et al., "Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu.," *Int. J. Epidemiol.*, vol. 44, no. 4, pp. 1137–47, Aug. 2015. <https://doi.org/10.1093/ije/dyt268>
3. C. Offord, "Learning from Iceland's Model for Genetic Research. The Scandinavian island's unique combination of genetic homogeneity, genealogical tradition, and high participation in research make it a prime location for discovery and validation of drug targets.," *The Scientist*, 2017. [Online]. Available: <https://www.the-scientist.com/bio-business/learning-from-icelands-model-for-genetic-research-31435>. [Accessed: 01-Jun-2017]. [MISSING]
4. NordForsk, "Nordic biobanks and registers A basis for innovative research on health and welfare," 2017.
5. D. Dolan, J. Grainger, N. MacCallum, D. Creatura, J. Forrester, and S. Shiller, "The Institute for Clinical Evaluative Sciences: 20 Years and Counting," *Healthc. Q.*, vol. 15, no. 4, pp. 19–21, Oct. 2012. <https://doi.org/10.12927/hcq.2012.23194>
6. A. Young and F. Flack, "Recent trends in the use of linked data in Australia.," *Aust. Health Rev.*, vol. 42, no. 5, pp. 584–590, Sep. 2018. <https://doi.org/10.1071/AH18014>
7. B. Haarbrandt et al., "HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries.," *Methods Inf. Med.*, vol. 57, no. S 01, pp. e66–e81, 2018. <https://doi.org/10.3414/ME18-02-0002>
8. The National Patient Centered Clinical Research Network: PCORNET [Online]. Available: <https://pcornet.org/>. [Accessed: 21-Nov-2018].
9. J. M. Gaziano et al., "Million Veteran Program: A mega-biobank to study genetic influences on health and disease.," *J. Clin. Epidemiol.*, vol. 70, pp. 214–23, Feb. 2016. <https://doi.org/10.1016/j.jclinepi.2015.09.016>
10. "https://allofus.nih.gov/about/all-us-research-program-protocol." [Online]. Available: <https://allofus.nih.gov/about/all-us-research-program-protocol>. [Accessed: 21-Nov-2018].
11. J. Pell, J. Valentine, and H. Inskip, "One in 30 people in the UK take part in cohort studies," *The Lancet*. 2014. [https://doi.org/10.1016/S0140-6736\(14\)60412-8](https://doi.org/10.1016/S0140-6736(14)60412-8)
12. Secretary of State for Health announces ambition to map 5 million genomes. 2 October 2018. [Online]. Available: <https://www.gov.uk/government/news/m>



- [att-hancock-announces-ambition-to-map-5-million-genomes](#). [Accessed: 21-Nov-2018].
13. J. Adams, T. Loach, and M. Szomszor, "Interdisciplinary research: methodologies for identification and assessment. Do we know what we are measuring?," London Digit. Sci., 2016.
  14. Enhancing the Effectiveness of Team Science. National Academies Press, 2015.
  15. A. Casadevall and F. C. Fang, "Revolutionary science," *MBio*, vol. 7, no. 2, pp. 1–6, 2016.
  16. T. F. C. Institute, "Discovery Without Boundaries," 2013. <https://doi.org/10.1128/mBio.00158-16>.
  17. K. Bhaskaran, I. Douglas, H. Forbes, I. Dos-Santos-Silva, D. A. Leon, and L. Smeeth, "Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults," *Lancet*, vol. 384, no. 9945, pp. 755–65, Aug. 2014. [https://doi.org/10.1016/S0140-6736\(14\)60892-8](https://doi.org/10.1016/S0140-6736(14)60892-8).
  18. S.-C. Chung et al., "Acute myocardial infarction: a comparison of short-term survival in national outcome registries in Sweden and the UK.," *Lancet* (London, England), vol. 383, no. 9925, pp. 1305–12, Apr. 2014. [https://doi.org/10.1016/S0140-6736\(13\)62070-X](https://doi.org/10.1016/S0140-6736(13)62070-X)
  19. E. Rapsomaniki et al., "Blood pressure and incidence of twelve cardiovascular diseases: Lifetime risks, healthy life-years lost, and age-specific associations in 1.25 million people," *Lancet*, vol. 383, no. 9932, pp. 1899–1911, 2014. [https://doi.org/10.1016/S0140-6736\(14\)60685-1](https://doi.org/10.1016/S0140-6736(14)60685-1).
  20. A. D. Shah, J. W. Bartlett, J. Carpenter, O. Nicholas, and H. Hemingway, "Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study," *Am. J. Epidemiol.*, vol. 179, no. 6, pp. 764–74, Mar. 2014. <https://doi.org/10.1093/aje/kwt312>
  21. N. Conrad et al., "Temporal trends and patterns in heart failure incidence: A population-based study of 4 million individuals," *Lancet*, vol. 6736, no. 17, 2017. [https://doi.org/10.1016/S0140-6736\(17\)32520-5](https://doi.org/10.1016/S0140-6736(17)32520-5)
  22. N. Freemantle et al., "Increased mortality associated with weekend hospital admission: a case for expanded seven day services?," *Br Med J*, vol. 351, p. h4596, 2015. <https://doi.org/10.1136/bmj.h4596>
  23. Denaxas S et al. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. *Journal of the American Medical Informatics Association*. 2019 Dec 1;26(12):1545–59. <https://caliberresearch.org/portal> <https://doi.org/10.1093/jamia/ocz105>
  24. K. I. Morley et al., "Defining Disease Phenotypes Using National Linked Electronic Health Records: A Case Study of Atrial Fibrillation," *PLoS One*, vol. 9, no. 11, p. e110900, 2014. <https://doi.org/10.1371/journal.pone.0110900>
  25. S.-M. Zhou et al., "Defining Disease Phenotypes in Primary Care Electronic Health Records by a Machine Learning Approach: A Case Study in Identifying Rheumatoid Arthritis," *PLoS One*, vol. 11, no. 5, p. e0154515, May 2016. <https://doi.org/10.1371/journal.pone.0154515>
  26. ClinicalCodes.org. [Online]. Available: [www.clinicalcodes.org](http://www.clinicalcodes.org). [Accessed: 21-Nov-2018].
  27. R. Jackson et al., "CogStack -Experiences of Deploying Integrated Information Retrieval and Extraction Services in a Large National Health Service Foundation Trust Hospital." <https://doi.org/10.1101/123299>
  28. European Federation for Medical Informatics Association (EFMI) and Farr Institute: Informatics for Health, Manchester 2017 <https://healthmanagement.org/c/it/event/informatics-for-health-congress-2017> [Accessed: 21- Nov-2018].
  29. J. Ainsworth and I. Buchan, "Combining Health Data Uses to Ignite Health System Learning.," *Methods Inf. Med.*, vol. 54, no. 6, pp. 479–87, 2015. <https://doi.org/10.3414/ME15-01-0064>
  30. NHS England, "Local Government Association. Local Health and Care Record Exemplars," 2018.
  31. J. Temperton, "NHS care.data scheme closed after years of controversy," *Wired*, 2016.
  32. W. W. Stead, "Clinical Implications and Challenges of Artificial Intelligence and Deep Learning.," *JAMA*, vol. 320, no. 11, pp. 1107–1108, 2018. <https://doi.org/10.1001/jama.2018.11029>
  33. HDR UK, "One Institute Strategy 2019/20."
  34. C. S. Wagner, T. A. Whetsell, and L. Leydesdorff, "Growth of international collaboration in science: revisiting six specialties," *Scientometrics*, vol. 110, no. 3, pp. 1633–1652, Mar. 2017. <https://doi.org/10.1007/s11192-016-2230-9>
  35. M. Schmidt, S. A. J. Schmidt, J. L. Sandegaard, V. Ehrenstein, L. Pedersen, and H. T. Sørensen, "The Danish National patient registry: A review of content, data quality, and research potential," *Clinical Epidemiology*, vol. 7. Dove Medical Press Ltd, pp. 449–490, 17-Nov-2015. <https://doi.org/10.2147/CLEP.S91125>
  36. S. and T. C. The House of Lords, "Genomic Medicine," vol. I:Report, 2009.
  37. S. Denaxas et al., "UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER," *J. Am. Med. Informatics Assoc.*, vol. 0, no. 0, pp. 1–15, 2019. <https://doi.org/10.1093/jamia/ocz105>.

38. V. Kuan et al., "A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service," *Lancet Digit. Heal.*, vol. 7500, no. 19, pp. 1–15, 2019. [https://doi.org/10.1016/S2589-7500\(19\)30012-3](https://doi.org/10.1016/S2589-7500(19)30012-3)
39. HEFC, "Panel criteria and working methods - REF01 Part 2B," no. January, 2012.
40. Wellcome Trust, Understanding Patient Data <https://understandingpatientdata.org.uk> [Accessed: 21-Nov-2018].
41. E. I. Benchimol et al., "The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement," *PLOS Med.*, vol. 12, no. 10, p. e1001885, Oct. 2015. <https://doi.org/10.1371/journal.pmed.1001885>
42. A. D. Shah et al., "Type 2 diabetes and incidence of cardiovascular diseases: a cohort study in 1.9 million people," *Lancet Diabetes Endocrinol.*, vol. 3, pp. 105–113, 2015. [https://doi.org/10.1016/S2213-8587\(14\)70219-0](https://doi.org/10.1016/S2213-8587(14)70219-0)
43. E. Herrett et al., "Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study," *BMJ*, vol. 346, p. f2350, Jan. 2013. <https://doi.org/10.1136/bmj.f2350>
44. S. M. Langan et al., "Herpes Zoster Vaccine Effectiveness against Incident Herpes Zoster and Post-herpetic Neuralgia in an Older US Population: A Cohort Study," *PLoS Med.*, vol. 10, no. 4, p. e1001420, Apr. 2013. <https://doi.org/10.1371/journal.pmed.1001420>
45. T. P. van Staa et al., "The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: Evaluations of two exemplar trials," 2014. <https://doi.org/10.3310/hta18430>
46. B. D. B. Bray et al., "Weekly variation in health-care quality by day and time of admission: a nationwide, registry-based, prospective cohort study of acute stroke care," *Lancet*, vol. 0, no. 0, pp. h4652-668, May 2016. [https://doi.org/10.1016/S0140-6736\(16\)30443-3](https://doi.org/10.1016/S0140-6736(16)30443-3)
47. V. M. Narasimhan et al., "Health and population effects of rare gene knockouts in adult humans with related parents.," *Science*, vol. 8624, no. March, pp. 1–8, Mar. 2016. <https://doi.org/10.1126/science.aac8624>
48. K. H. Jones, D. V. Ford, S. Ellwood-Thompson, and R. A. Lyons, "The UK Secure eResearch Platform for public health research: a case study," *Lancet*, vol. 388, p. S62, Nov. 2016. [https://doi.org/10.1016/S0140-6736\(16\)32298-X](https://doi.org/10.1016/S0140-6736(16)32298-X)
49. Bauermeister S. et al. The Dementias Platform UK (DPUK) Data Portal [Internet]. Neuroscience; 2019 Mar. Available <https://doi.org/10.1093/jamia/ocz105>
50. "https://www.dementiasplatform.uk/news/uk-wide-research-programme-to-investigate-dementia-in-stroke-patients." [Online]. Available: <https://www.dementiasplatform.uk/news/uk-wide-research-programme-to-investigate-dementia-in-stroke-patients>. [Accessed: 19-Sep-2019].
51. SAIL Databank [Online]. Available: [www.saildatabank.com](http://www.saildatabank.com). [Accessed: 28-Nov-2018].
52. K. H. Jones et al., "A case study of the secure anonymous information linkage (SAIL) gateway: A privacy-protecting remote access system for health-related research and evaluation," *J. Biomed. Inform.*, vol. 50, pp. 196–204, 2014. <https://doi.org/10.1016/j.jbi.2014.01.003>
53. Health eResearch Centre, Manchester University. [Online]. Available: [www.herc.ac.uk/tre](http://www.herc.ac.uk/tre). [Accessed: 18-Nov-2018].
54. The electronic Data Research and Innovation Service (eDRIS) part of the Information Services Division. [Online]. Available: [www.isdscotland.org/products-and-services/eDRIS](http://www.isdscotland.org/products-and-services/eDRIS). [Accessed: 28-Nov-2018].
55. Connected Health Cities Impact Summary: 2016-2020 Delivering trustworthy Data Driven improvements in care for our patient population Northern Health Science Alliance and CCH, 2020. [Online]. Available: [www.connectedhealthcities.org](http://www.connectedhealthcities.org). [Accessed: 28-Feb-2020].
56. Medical Research Council Funded Cloud Compute: the eMEDLAB <https://mrc.ukri.org/news/browse/mrc-funded-supercomputer-emedlab-is-an-invaluable-to-researchers-using-big-data/> [Accessed: 28-Nov-2018].
57. Medical Research Council Cloud Infrastructure for Microbial Bioinformatics [Online]. Available: <https://www.climb.ac.uk/>. [Accessed: 28-Nov-2018].
58. Joint Information Systems Committee: Safe Share [Online]. Available: <https://www.jisc.ac.uk/safe-share>. [Accessed: 28-Nov-2018].
59. J. Aboab et al., "A 'datathon' model to support cross-disciplinary collaboration," *Science Translational Medicine*, vol. 8, no. 333. American Association for the Advancement of Science, 06-Apr-2016. <https://doi.org/10.1126/scitranslmed.aad9072>
60. R. J. Harris et al., "High-throughput image labeling and quality control for clinical trials using machine learning," *Int. J. Clin. Trials*, vol. 5, no. 4, p. 161, Oct. 2018. <https://doi.org/10.18203/2349-3249.ijct20184398>

61. D. A. Springate et al., "ClinicalCodes: An online clinical codes repository to improve the validity and reproducibility of research using electronic medical records," *PLoS One*, vol. 9, no. 6, Jun. 2014. <https://doi.org/10.1371/journal.pone.0099825>
62. S. Sawhney and S. D. Fraser, "Epidemiology of AKI: Utilizing Large Databases to Determine the Burden of AKI," *Advances in Chronic Kidney Disease*, vol. 24, no. 4. W.B. Saunders, pp. 194–204, 01-Jul-2017. <https://doi.org/10.1053/j.ackd.2017.05.001>
63. S. Sawhney et al., "Acute kidney injury in the UK: A replication cohort study of the variation across three regional populations," *BMJ Open*, vol. 8, no. 6, Jun. 2018. <https://doi.org/10.1136/bmjopen-2017-019435>
64. B. Fonferko-Shadrach et al., "Validating epilepsy diagnoses in routinely collected data," *Seizure*, vol. 52, pp. 195–198, Nov. 2017. <https://doi.org/10.1016/j.seizure.2017.10.008>
65. R. Jackson et al., "CogStack - Experiences Of Deploying Integrated Information Retrieval And Extraction Services In A Large National Health Service Foundation Trust Hospital," *bioRxiv*, 2017. <https://doi.org/10.1186/s12911-018-0623-9>
66. H. Wu et al., "SemEHR: A general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research\*," *J. Am. Med. Informatics Assoc.*, vol. 0, no. February, pp. 1–9, 2018. <https://doi.org/10.1093/jamia/ocx160>.
67. S. Alvarez-Madrado, S. McTaggart, C. Nangle, E. Nicholson, and M. Bennie, "Data resource profile: The Scottish national prescribing information System (PIS)," *International Journal of Epidemiology*, vol. 45, no. 3. Oxford University Press, p. 714F–715F, 2016. <https://doi.org/10.1093/ije/dyw060>
68. S. McTaggart, C. Nangle, J. Caldwell, S. Alvarez-Madrado, H. Colhoun, and M. Bennie, "Use of text-mining methods to improve efficiency in the calculation of drug exposure to support pharmacoepidemiology studies," *Int. J. Epidemiol.*, vol. 47, no. 2, pp. 617–624, Apr. 2018. <https://doi.org/10.1093/ije/dyx264>
69. RADAR-CNS: Remote Assessment of Disease and Relapse – Central Nervous System, "<https://www.radar-cns.org/>." <https://doi.org/10.1093/ije/dyx264>
70. A. Story et al., "Smartphone-enabled video-observed versus directly observed treatment for tuberculosis: a multicentre, analyst-blinded, randomised, controlled superiority trial," *Lancet*, vol. 393, no. 10177, pp. 1216–1224, Mar. 2019. [https://doi.org/10.1016/S0140-6736\(18\)32993-3](https://doi.org/10.1016/S0140-6736(18)32993-3)
71. Centers for Disease Control and Prevention: Shingles (herpes zoster) [Online]. Available: <https://www.cdc.gov/vaccines/acip/recs/grade/herpes-zoster.html>. [Accessed: 28-Nov-2018]. 72.
72. World Health Organisation: background document on immunisation with herpes zoster vaccines. [Online]. Available: [https://www.who.int/immunization/sage/meetings/2014/april/2\\_Background\\_document\\_Herpes\\_Zoster.pdf](https://www.who.int/immunization/sage/meetings/2014/april/2_Background_document_Herpes_Zoster.pdf). [Accessed: 28-Nov-2019]
73. C. A. Marwick et al., "A multifaceted intervention to improve sepsis management in general hospital wards with evaluation using segmented regression of interrupted time series," *BMJ Qual. Saf.*, vol. 23, no. 12, p. e2, Dec. 2014. <https://doi.org/10.1136/bmjqs-2013-002176>
74. "[http://www.sehd.scot.nhs.uk/publications/DC20151126nti\\_15-16\\_Full\\_Document\\_FINAL.pdf](http://www.sehd.scot.nhs.uk/publications/DC20151126nti_15-16_Full_Document_FINAL.pdf)."
75. D. J. McLernon, A. Maheshwari, A. J. Lee, and S. Bhattacharya, "Cumulative live birth rates after one or more complete cycles of IVF: A population-based study of linked cycle data from 178 898 women," *Hum. Reprod.*, vol. 31, no. 3, pp. 572–581, 2016. <https://doi.org/10.1093/humrep/dev336>
76. Scottish Government National Infertility Group 2016. ISBN: 978-1-78652-157-6 "<http://www.gov.scot/Resource/0050/00501403.pdf>."
77. R. W. Aldridge et al., "Tuberculosis in migrants moving from high-incidence to low-incidence countries: a population-based cohort study of 519 955 migrants screened before entry to England, Wales, and Northern Ireland," *Lancet*, vol. 388, no. 10059, pp. 2510–2518, Nov. 2016. [https://doi.org/10.1016/S0140-6736\(16\)31008-X](https://doi.org/10.1016/S0140-6736(16)31008-X).
78. "[https://www.acc.org/~media/Non-Clinical/Files-PDFs-Excel-MS-Word-etc/Guidelines/2017/Guidelines\\_Made\\_Simple\\_2017\\_HBP.pdf](https://www.acc.org/~media/Non-Clinical/Files-PDFs-Excel-MS-Word-etc/Guidelines/2017/Guidelines_Made_Simple_2017_HBP.pdf)."
79. V. Allan et al., "Net clinical benefit of warfarin in individuals with atrial fibrillation across stroke risk and across primary and secondary care," *Heart*, vol. 103, no. 3, pp. 210–218, 2017. <https://doi.org/10.1136/heartjnl-2016-309910>
80. A. Timmis et al., "Prolonged dual antiplatelet therapy in stable coronary disease: comparative observational study of benefits and harms in unselected versus trial populations," *BMJ*, vol. 353, 2016. <https://doi.org/10.1136/bmj.i3163>
81. Ticagrelor for preventing atherothrombotic events after myocardial infarction National Institute of health and Clinical Excellence. Technology appraisal guidance [TA420] Published date: 14 December 2016 [www.nice.org.uk/guidance/TA420/chapter/1-Recommendations80](http://www.nice.org.uk/guidance/TA420/chapter/1-Recommendations80)." [Online].