## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>02</td>
</tr>
<tr>
<td>Foreword</td>
<td>03</td>
</tr>
<tr>
<td>Executive Summary</td>
<td>05</td>
</tr>
<tr>
<td>Research Highlights 2015</td>
<td>09</td>
</tr>
<tr>
<td>- Substantive Research</td>
<td>09</td>
</tr>
<tr>
<td>- Underpinning Methodology</td>
<td>11</td>
</tr>
<tr>
<td>- Research Centred Around Disease Themes</td>
<td>12</td>
</tr>
<tr>
<td>Enabling Data Sets and Developing New Infrastructure, Methods, Technologies and Standards for Research</td>
<td>13</td>
</tr>
<tr>
<td>- Innovative Governance</td>
<td>13</td>
</tr>
<tr>
<td>- Trusted Research Environments</td>
<td>14</td>
</tr>
<tr>
<td>- Enabling New Access for Researchers to EHR Data Sets</td>
<td>15</td>
</tr>
<tr>
<td>- UK SeRP and Distributed Access to Consented Cohort Studies</td>
<td>15</td>
</tr>
<tr>
<td>- Linkage to Consented Cohorts to EHR Sources</td>
<td>15</td>
</tr>
<tr>
<td>- Methods and Tools</td>
<td>16</td>
</tr>
<tr>
<td>Capacity Building</td>
<td>17</td>
</tr>
<tr>
<td>Patient and Public Involvement and Engagement</td>
<td>19</td>
</tr>
<tr>
<td>Knowledge Transfer</td>
<td>21</td>
</tr>
<tr>
<td>Partnerships and Collaborations</td>
<td>23</td>
</tr>
<tr>
<td>Farr Institute eHealth Informatics Centre Reports</td>
<td>25</td>
</tr>
<tr>
<td>- Farr CIPHER</td>
<td>25</td>
</tr>
<tr>
<td>- Farr HeRC</td>
<td>27</td>
</tr>
<tr>
<td>- Farr London</td>
<td>29</td>
</tr>
<tr>
<td>- Farr Scotland</td>
<td>31</td>
</tr>
<tr>
<td>Appendices</td>
<td>33</td>
</tr>
<tr>
<td>The Farr Institute Publications 2015</td>
<td>34</td>
</tr>
<tr>
<td>Glossary</td>
<td>38</td>
</tr>
</tbody>
</table>
It is a great pleasure to present The Farr Institute Annual Report 2015/16.

Following his Eight Great Technologies Science speech in January 2013, David Willetts (the then Universities and Science Minister), announced funding for The Farr Institute in July 2013. The strategic aim was to strengthen the MRC coordinated, ten partner £17.4m award that had created four e-Health Informatics Research Centres (eHIRCs) in the UK in 2012. This assembled 21 UK Universities and two MRC Units to create four virtual research centres, hosted by the University of Manchester, University of Dundee, Swansea University and University College London (UCL). The eHIRCs had started to create much-needed multi-disciplinary informatics capacity (25 new senior academic posts) and were harnessing linked health data for discovery science, experimental medicine and data-intensive innovations in healthcare. A £400K per annum award created a Network across the Centres to support UK-wide communities of best practice, coordinated training, sharing of methodology and concerted engagement with the public, industry and the NHS.

The additional £20m MRC capital award provided e-infrastructure and physical buildings to lay the foundations of a more formal national distributed research facility across the Centres. The Farr Institute is named after William Farr (1807-83), who was a pioneer in early computation (with Babbage’s Difference Engine) of health data, underpinning public health reforms.

The additional £20m MRC capital award provided e-infrastructure and physical buildings to lay the foundations of a more formal national distributed research facility across the Centres. The Farr Institute is named after William Farr (1807-83), who was a pioneer in early computation (with Babbage’s Difference Engine) of health data, underpinning public health reforms.

Our vision for The Farr Institute is to unleash a bigger scale of public-serving research across the UK’s health and health-related data assets, particularly those that can be linked for defined populations. UK data range from healthcare records to biological, social, economic and environmental observations. The bold aim of The Farr Institute is to establish the UK as the world leader in health data science research from scientific discovery to the enhancement of patient and public health.

There is an urgent need to move beyond the constraints of a classical ‘research collaboration’ that the original investments funded. The clear opportunity is to leverage the scale of the UK’s diverse populations by creating an agile national ecosystem of universities, NHS, private companies and other stakeholder organisations. To achieve this will require a massively scalable apparatus that can enable large, multi-centre, deeply-phenotyped studies at local, regional and national levels, where the focus is the alignment of all available biomedical data per individual, with the ability to link health and healthcare history, genes, other biological information, environments, and lifestyles in a trustworthy way.

The apparatus and ecosystem that The Farr Institute is building is a vital first step for conducting ‘big data’ longitudinal studies: enabling a bigger return from larger investments such as UK Biobank and Genomics England; uncovering new ‘phenomarkers’ for precision medicine; and enabling rapid translation of science into more data-responsive healthcare, public health and economic developments.

The Farr Institute brand has rapidly become established internationally. There is no other national effort with similar aims, however, there is much to be done to realise this bold, ambitious and enterprising agenda. We are committed to working in partnership with colleagues across the UK to grasp this opportunity.

This report marks the achievements of the eHIRCs against their original objectives and contracts (of February 2012). We also describe how we have focussed the objectives of The Farr Institute; integrated existing activities; made a series of strategic recruitments; created outstanding inter-disciplinary research facilities in Swansea, London, Manchester, Liverpool, Dundee and Edinburgh; and developed the Institute’s digital infrastructure.

Finally, we document how we are committed to working in partnership. The UK landscape will have to adapt if we are to be the come-to-place for health, biomedical informatics and discovery science research that exploits the nation’s leadership position in life sciences, informatics and genomics.
It is a very complex landscape and the cross-sectoral inter-dependencies are at times daunting. We report on joint working with many partners; a non-exhaustive list includes the Health and Social Care Information Centre (HSCIC); NHS England, NHS Scotland, NHS Wales; Northern Health Science Alliance (NHSA); Research Councils UK (RCUK); Clinical Practice Research Datalink (CPRD); research charities including British Heart Foundations (BHF) and Cancer Research UK (CRUK); Innovate UK; Joint Information Systems Committee (Jisc), Association of British Pharmaceutical Industries (ABPI); Academy of Medical Sciences; Administrative Data Research Network (ADRN); Alan Turing Institute and the UK Health Data Analytics Network. The Institute is also working internationally with: The International Medical Informatics Association (IMIA); World Health Organisation (WHO); the Global Alliance for Genomic Health (GA4GH); National Institutes of Health (NIH); National Science Foundation (NSF); and with large-scale health data initiatives in the US, Canada, Netherlands, Sweden, Denmark, Australia, New Zealand, Canada, Singapore, China, Malaysia and South America.

Over the past three years, the Institute has started to put the UK on the global map of informatics and data science for health. We now seek to consolidate and extend this work and to work in partnership to realise the research value of the UK’s considerable health data assets. This reflects the ambitious, far-reaching recommendations of our International Advisory Board, starting with the appointment of Dr Graham Spittle, our Strategic Advisor.

It has been a great privilege to have been invited to lead on the creation and establishment of The Farr Institute. This however has been an incredible team effort and we are enormously grateful to the outstanding scientific, clinical, academic and administrative teams across the UK who have worked tirelessly and with massive commitment to make the achievements described in this document possible.

Iain Buchan is Clinical Professor in Public Health Informatics at the University of Manchester, where he founded and leads the Centre for Health Informatics.

Harry Hemingway is Professor of Clinical Epidemiology at University College London and Director of the Farr Institute in London.

Ronan Lyons is Professor of Public Health, Swansea University and Honorary Consultant in Public Health with Public Health Wales.

Andrew Morris is Professor of Medicine, Director of the Usher Institute of Population Health Sciences and Informatics and Vice Principal of Data Science at the University of Edinburgh.
The vision of The Farr Institute is to unleash a larger scale of public-serving research across the UK’s health and health-related data assets, particularly those that can be linked for defined populations. Our aim is to establish the UK as a world leader in health data science research from discovery science to the enhancement of patient and public health. The Institute was proposed by MRC in July 2013, in the belief that the UK has a unique global opportunity, built on:

- A rich tradition of UK excellence and innovation in using routine health (care) data for pioneering research stretching back over 150 years to when William Farr systematised the analysis of health data in England and Wales, with enduring impacts on public health and policy-making.
- Existing strengths in mathematics, biostatistics, epidemiology, genomics, bioinformatics, and computer science, but with limited health and biomedical informatics capacity needed for more integrative health data science.
- Academic excellence in a range of disciplines across the eHIRC Centres which could be readily complemented by forging collaborations with other centres of excellence across the UK and internationally.
- Internationally-recognised excellence in applied research with health data including our longstanding investments in UK birth and case cohorts, greatly enhanced by the existence of the NHS and allied public services where patients/citizens can be tracked across whole systems.
- The international commitment to precision medicine, that facilitates translation of basic scientific progress into clinical impact through effective research, underpinned by informatics, deep phenotyping and advanced analytics.

Delivery will require strong UK leadership, greater interoperability through widespread use of standards, new methods to support the integration of heterogeneous structured and unstructured data, and an inclusive approach to forge collaboration with other initiatives and bodies [e.g. MRC Medical Bioinformatics Centres, Genomics England (GeL), National Institute for Health Research (NIHR), Health and Social Care Information Centre (HSCIC)]. It will also require the development of enduring partnerships among multiple academic, NHS and industry bodies and a rapid increase in the UK skill base. These are all critical factors in our quest to perform transformational science at scale and drive economic growth, whilst maintaining the trust of the public in everything we do.

At the first meeting of The Farr Institute International Advisory Board in December 2014, chaired by Professor Nancy Pedersen (Karolinska Sweden) five major recommendations were made to (i) Communicate a clear vision of the identity, role, remit, objectives and added value of The Farr Institute (the basis of this report);

(ii) Unify and integrate the research, education and expertise within each of the four Farr Centres (as reflected in the UK emphasis of this report); (iii) Deliver 2-3 cross-centre demonstrator projects which show the added value contributed by The Farr; (achieved – see research highlights); (iv) Take a national lead in placing public/patient involvement and advocacy at the heart of The Farr Institute; and, (v) Become a leader on best practice in Governance relating to health informatics research, working with similar organisations internationally to address governance issues that hinder global research.

The Farr Institute (Box 1) sets out The Farr Institute objectives for 2015-16. In this 2015/16 Annual Report the progress and impact across these six domains is described.

---

**Box 1. 2015-16 Objectives**

In 2015/16 The Farr Institute aligned its purpose and function and agreed six objectives designed to provide a powerful foundation for UK informatics and health data science;

1. Perform pioneering multi-disciplinary research with large and complex health-related data, including healthcare records, biological, social, and environmental data.
2. Enable new datasets and develop new infrastructure, methods, technologies, and standards for such research.
3. Develop skills, talent and expertise in individuals and research communities for collaborative working.
4. Work with the owners and controllers of data to support the safe use of patient information for medical research across the UK, championing data protection, confidentiality and privacy.
5. Engage with the public to demonstrate the benefits of using health data in research and to encourage the support of secure and trusted access to patient information.
6. Bring together government, public sector, academia and industry to foster relationships and establish best practices for innovation, discovery and impact in health data science.
Delivery against objectives

Research. The Farr Institute has published 74 papers directly from the eHIRC funding (in a range of journals including Science, Lancet, BMJ, PLOSMed), across a breadth of research domains including public health, discovery science, precision medicine, learning health systems and ethical and social issues in relation to health records research (Appendix 1). A further 328 papers with other colleagues have been enabled by Farr-funded scientists and engineers.

The Farr Institute has established seven UK health data science research groups in Asthma (in collaboration with Asthma UK); Mental Health (with MQ); Renal Disease; Cancer; Drug Safety; Linking UK cohorts and Primary Care. Farr Investigators have leveraged an additional £68M in research funding since 2013 in areas including: cancer; dementia; industry engagement.

Enabling new data sets. Farr has worked with data controllers to start to provision new data sets for research including: national prescribing (100M per annum) and national imaging (23M since 2009) for the 5M population of Scotland; critical care data, laboratory data and the referrals dataset for the 3M population of in Wales; the National Neonatal Research Database and National Hip Fracture Database. The Farr Institute is committed to support NIHR and ELIXIR (the EU infrastructure for biological information) to create appropriate data discovery services for data assets to promote their availability on behalf of data controllers (whether NHS, Academic or other bodies).

Capacity building. The Farr Institute has created a UK wide doctoral training support programme for more than 70 PhD students and supported them with summer schools and symposia. Taught postgraduate masters courses, and online distance learning courses in health data science and technology have been developed in six member universities, matriculating over 400 students in 2015/16. 17 new faculty positions have been created across Farr Institute partners in 2015/16. UK contributions to the top five journals and top three international conferences in health and biomedical informatics have increased substantially.

Trusted Research Environments (TREs), policy and governance. The Farr Institute subscribes to a model of interoperability delivered by a federated network of TREs that enables the identification and retrieval of all data that pertain to individual health. This requires a data sharing architecture that is capacious enough for all relevant data types and that enables patient and institutional autonomy to be respected. This is a grand challenge for the UK. In 2015, The Farr Institute received ISO27001 accreditation for five TREs across the UK in Swansea, London, Dundee, Edinburgh and Glasgow. Manchester is in the process of gaining accreditation. The Farr Institute has demonstrated thought leadership by influencing Government policy in Wales and Scotland to create single information governance sign off and scrutiny for cross-sectoral health data science projects. In addition, Farr-supported research led to Government policy to create a network of “Connected Health Cities” across the 17M population for North England enabling connections between NHS, local government (including social care) and patient-collected data in the quest to create the world’s first true learning health systems in the UK.

Thought leadership, public engagement and advocacy. Senior Farr Investigators have been invited to deliver 45 plenary lectures during 2015/16 in the UK and 17 internationally. The Farr Institute has hosted high level day-long delegations from NHS England (Sir Malcolm Grant, Chairman), Innovate UK (Ruth McKearnan, CEO), HSCIC (Andy Williams, CEO), Office for Life Sciences (George Freeman MP, Minister for Life Sciences) and the Scottish Council for Economic Advisors (Nicola Sturgeon, First Minister, and John Swinney, Deputy First Minister). Farr Investigators have contributed to the
Executive Summary

Caldicott Review, Academy of Medical Sciences reviews of Team Science and Public Health, the Nuffield Council on Bioethics review of ‘The collection, linking and use of data in biomedical research and health care’, and the multi-funder Expert Advisory Group on Data Access (EAGDA) hosted by the Wellcome Trust that has published several influential reports including the public perception of data use for commercial purposes (March 2016). The Farr Institute has championed the #datasaveslives campaign which has received 17,000 mentions on Twitter and has been promoted by organisations such as Cancer Research UK (CRUK), Public Health England (PHE) and the government backed Empower data4health campaign.

The website attracted nearly 50,000 unique visitors over the past year and Twitter accounts have gained over 4,000 followers with a peak in activity of 43k impressions during The Farr Institute International Conference 2015. Investigators and projects from The Farr Institute have received eight mentions in the national press including three full length BBC radio interviews. The Farr Institute Directors were signatories on the letter sent to the Times on behalf of research institutes across Europe warning of the risk to research if the proposed change to the EU Data Protection Regulation had been adopted.

Industrial/academic collaborations. The Farr Institute, working with One Nucleus, the Catapults, Innovate UK and UK Trade and Investment (UKTI) has created a network of links to companies globally, across the pharmaceutical, digital health, biotechnology, informatics and data science industry verticals. Currently 15 active cross-centre projects are being developed (some are currently subject to non-disclosure agreements). Three have led to formal contracts.

The Farr Institute was the topic of a plenary address at several high profile national/international forums including UKTI, Drug Information Association (DIA) European Conference, Research Councils UK (RCUK) Research Innovate Grow event at Westminster and Expo Milan.

Opportunities

Through its leadership roles Farr investigators are able to seize rapidly emerging opportunities in health data science. Farr investigators lead or collaborate with initiatives in Health North’s Connected Health Cities, MRC Stratified Medicine, MRC Medical Bioinformatics Awards, Horizon 2020, Innovative Medicines Initiative (IMI) Big Data for Better Outcomes, Genomics England, Alan Turing Institute. As one example, the NIHR Biomedical Research Centres (current competition for new applications and renewals 2016) are likely to make substantial investments relevant to hospital informatics. In Wales, five new Health and Care Research Wales centres and units have new investments in data science; cancer, population health, renal, primary and emergency care, and mental health. Across North England from 2016-2019 there will be pilots of four “Connected Health Cities” each with a civic-centred data analytic facility drawing on data from HSCIC and local sources, leveraging expertise from The Farr Institute in informatics and statistics. A Connected Health Cities Hub will work closely with HSCIC and Department of Health to inform plans for scaling data analytics across multiple health systems, enabling deeper discovery science, more coordinated discovery science and more agile translation of research into algorithms and actionable care information.

In Scotland, The Farr Institute will be a founding partner of the recently launched Precision Medicine Ecosystem, that brings together each Scottish medical school, NHS Scotland and the national Innovation Centres, sponsored by the First Minister of Scotland.

Looking forward, our bold plans for 2016-2018 are guided by our objectives, achievements to date, challenges and opportunities. For example, we will facilitate the development of new UK-wide research programmes. We will also propose to establish new UK-wide working groups in hospital informatics/
analytics, genomics and electronic health record (EHR) integration for discovery science, precision medicine and intelligent trials. In developing Trusted Research Environments, we will work collegiately to ensure the successful delivery of the Joint Information Service Committee (Jisc) SafeShare project which could have broad impact across the UK scientific community and, at a regional level, work with the London AHSNs to create a TRE for research of local NHS data for London. This would be a big prize for the UK.

We are also proposing to launch a new Farr Institute Leadership programme, and Farr Institute Exchange Programme, providing the next generation of UK researchers mentorship and exposure to the brightest minds internationally. We are convinced that such a concerted effort is vital if the UK is to succeed in its quest to be the destination of choice for health data science.

This report demonstrates the value delivered as a result of the core investment in the Institute, distinguishing between achievements that would not have occurred and those which have been accelerated due to the Institute. It is recognised that the range of projects referenced herein is wider than those directly funded by Farr Institute & eHIRC grants. Funding from other sources is acknowledged in the text.
Objective 1: Perform pioneering inter-disciplinary research with large and complex health record, biological and other data.

The Farr Institute is foremost and primarily a research Institute. In 2015/16, The Farr Institute has demonstrated the ability to generate new knowledge through data science with healthcare and biological data combinations (Narasimhan et al. Science, 2016) to impactful health services research (Freemantle et al, BMJ, 2015), laying the foundations of learning health systems (Ainsworth & Buchan, Methods Inf. Med. 2015). The Farr Institute is demonstrating how the cutting edge use of rich, large scale NHS records is transforming science and healthcare. Here we provide highlights of the Institute’s research which illustrate cross-centre working, the scaling of data analytics across the UK, the novelty of methods and emerging signs of impact on health and healthcare. A comprehensive list of 2015/16 outputs (74 publications) is presented in Appendix 1.

Substantive Research

1. Discovering the real-world health relevance of ‘human knockouts’. In one of the first examples of English NHS data playing a central role in addressing a fundamental question of biological science, Farr investigators from multiple universities and international collaborators linked genome sequencing data in parentally related adults to life-long primary care health records. It was shown that loss-of-function gene knockouts were not associated with clinical consultation or prescription rates (Narasimhan et al. Science, 2016).

2. Precision medicine. The decision to prolong dual anti-platelet therapy in people with stable coronary disease requires individualised balancing of harms and benefits. EHRs in England were used to develop and validate prognostic risk models for both benefits and harms (bleeds) and demonstrated for each individual the net clinical benefit (Pasea ACC 2015, AZ-APOLLO programme).

3. Discovering and replicating novel associations at scale across the UK. LDL cholesterol is a marker of therapeutic efficacy for five classes of lipid lowering drug. Yet how it relates to common (heart failure, atrial fibrillation) and rare diseases (subarachnoid haemorrhage, intracerebral haemorrhage and ventricular arrhythmia) has not been evaluated in large scale observational cohorts. We showed using linked EHRs in about 500k people in England (discovery) and 700k people in Wales (replication) that there is a strong inverse association with heart failure onset (Moayeri, Brophy, unpublished).

4. Igniting care quality and outcome debates. The association of weekend admission and mortality, based on analysis of English Hospital Episode Statistics (Freemantle et al BMJ 2015; Roberts et al. Lancet, 2015), has contributed to a set of political and policy developments which have had a profound impact on a generation of junior doctors. The size of the “weekend effect” is disease dependent. Farr Institute research has also bust the myth that (incentivised) primary care quality predicts premature mortality in England (Kontopantelis et al. BMJ 2015).

5. Putting NHS care and outcomes on the international stage. Farr investigators have demonstrated not only poorer survival following myocardial infarction in the UK compared with Sweden, but with higher levels of between hospital variation in care and 30-day mortality. This suggests important opportunities to improve survival in the UK (Chung et al. BMJ 2015).

6. Triggering international debate from UK linked data studies. Using National Institute for Cardiovascular Outcomes Research (NICOR) linked data on percutaneous coronary interventions and mortality, The Farr Institute showed not only that radial artery access conferred better survival than the traditional femoral route, but also that regional variation in uptake produced inequitable care (Mamas et al. Circulation 2016).

7. Supporting clinical decisions. Much research has been focussed on areas from ‘big data’ to new knowledge to actionable information. In Acute Kidney Injury NHS decision support collaborations with Google DeepMind have been announced and the NHS England ‘Think Kidneys’ programme cites the algorithm development of The Farr Institute. (Sawhney et al. Nephrol Dial Transplant., 2015).

9. Guiding care with mobile apps and frequently reported patient-driven outcomes.
Farr researchers developed www.clintouch.com which now enables high-resolution, continuous longitudinal monitoring of mental health symptom and awareness-raising. This has been further developed to help counter limitations of care in severe mental illness; unreliable symptom reporting; patient disengagement from care planning; avoidable hospital admissions following late detection of relapse (Whelan et al. Stud Health Technol Inform, 2015).

10. Delivering new insights for public health.
The policy value of linking fuel poverty data (Home Energy Efficiency) with routinely-collected health data has been demonstrated. In doing this, many lessons have been learned about data quality, the challenges associated with the use of newly-linked data sets and the methods required to analyse linked, longitudinal data (Heaven and Lowes, 2013 summary at is.gd/OW5f65).
Underpinning Methodological Research

1. Building global consensus for engineering learning health systems.
   An entire journal edition was devoted to a landmark paper from The Farr Institute (Ainsworth & Buchan Meth. Inform. Med. 2015) alongside commenting papers invited from the international experts in learning health systems. The leading editorial from the former President of the American Medical Informatics Association carried the title “At last! A Working Model of a Data Ecosystem for Continuous Learning in the Evolving Health Noosphere”.

2. Advancing biostatistical methods for EHR data analysis.
   Causal inference: the common “ collider bias” explanation for the paradox of moderate obesity appearing to be a protective factor in cohorts of people with type 2 diabetes was proven through simulation and statistical reasoning to be inadequate (Sperrin et al. Epidemiology 2016 in press). Information refinement: a generalizable model-based approach was shown to enhance spatial information on disease prevalence in low-resource settings (Diggle & Giorgi J. Am. Stat. Assoc. 2016 in press).

3. Understanding selection and reporting biases with connected health technologies.
   A collaboration with Withings revealed different male vs. female profiles of typical smart-scale users and substantial correlation between observation patterns following weight-change (Sperrin et al JIMIR 2016).

4. Enabling computer-assisted scaling and reproducibility of research.
   The Research Object and e-Lab methodology has been adopted in the US Big Data to Knowledge programme and the international FAIR (Find, Access, Interoperate, Reuse) principles (Wilkinson et al. Nature: Scientific Data 2016). NSF/NIH invited a related keynote (Buchan) in a UK-US workshop.

5. Machine-learning epidemiologically important patterns.
   Farr Institute-developed machine learning and biostatistical methods continue to uncover important disease endotypes, particularly with complex longitudinal data (e.g. Custovic et al. J. Allergy Clin. Immunol. 2015). Similar computational thinking has delivered new methods for mining more accurate prescribing information from clinical text and codes in EHRs (Karystianis BMC Med. Inform. 2016).

6. Delivering replicable longitudinal phenotypic resolution at scale from English, Scottish and Welsh EHR data in UK Biobank.
   2015 saw the first release of genotypic information from UK Biobank to the research community. Farr investigators have led efforts to deliver computable, replicable EHR phenotypes at scale and these phenotypes were shown to replicate known GWAS associations. One example is the ascertainment and phenotyping of myocardial infarction in 500k UK Biobank participants. We demonstrate that using a disease registry (MINAP) to distinguish ST elevation MI (M1) from non-ST elevation MI (M2), compared with use of an admission ICD-10 code (M3), or self-report alone (M7). Figure 1 shows how different all-cause mortality is in prevalent cases defined in these different ways.

7. Enhancing the longitudinal resolution of phenomarkers from care records.
   Using linked primary and secondary care EHR data on around 2m people aged ≥ 30 years, and free from diagnosed cardiovascular disease, we used the CALIBER research platform to understand key cardiovascular outcomes and presentations (George et al. Circulation 2015). Between 1997 and 2010, nearly 115,000 people in this EHR cohort experienced an incident cardiovascular diagnosis. However, most (66%) were neither myocardial infarction nor ischemic stroke which most primary prevention studies focus on. Males and females have differing associations with the cardiovascular diseases with implications for risk prediction and management strategies. So EHR phenomarking needs to influence discovery science not just follow it.

8. Transatlantic data science on UK EHRs.
   The NIH/RCUK-funded workshop Bethesda 2016, built around three UK and three US data sources, Farr investigators presented the only EHR based resource – CALIBER. This is a resource of national EHR records across primary, secondary and tertiary care linked to disease registries (~14m registered patients). The computational methods and tools for querying, extracting and creating research cohorts for answering clinically meaningful questions has sparked collaborations with Harvard and Florida State University principally in machine learning.

9. Standardising EHR study reporting.
   Such standards are vital and have been influential in providing a window into the quality, transparency and replicability of research in other areas. The Farr Institute underpinned the RECORD statement landmark international

10. Linking data to understand the wider determinants of health. The Farr Institute is co-located with the ESRC Administrative Data Research Centres (ADRC), enabling data linkages between health and education (e.g. National Pupil Database to ALSPAC). This can reduce bias (e.g. due to missing IQ in observational studies) and help impute missing data. (Cornish et al. Int. J. Epidemiol. 2015).

In Scotland, The Farr Institute and ADRC have integrated their governance and e-infrastructure to improve efficiency, thus achieving economies of scale, avoiding duplication and better return on investment of RCUK funding (Pavis & Morris Public Health Pract. 2015), now supporting over 300 academic, NHS and industry projects.

11. Advancing the ethics and social understanding of health data uses. The concept of a “social license” was introduced by Farr researchers to explain public concern over the care.data plans to centralise multiple uses of NHS records in England (Carter P, Laurie G, et al. J Med Ethics, 2015). The social license explains the need to look beyond just legal and regulatory compliance and consider the different dimensions of patient acceptance and trust, and interpretation of data use as “public good”. These principles are being used in the Citizens’ Juries over health data uses in North England alongside new research into dynamic consent processes (Spencer et al. JMIR 2016).

Research Centered Around Disease Themes

In major disease areas The Farr Institute has brought together interdisciplinary UK collaborations to advance data-intensive research. Our approach has been to stimulate and coordinate strategic cross-centre funding applications as follows.

Asthma. The Farr Institute has joined forces with the 14-university Asthma UK Centre for Applied Research (AUKCAR, www.aukcar.ac.uk) (Director: A. Sheikh, Edinburgh and Co-I R. Lyons, Swansea). This collaboration is building an asthma-focussed national ‘learning health system’ through deep analysis of EHR data. Launched in Edinburgh in October 2015 with multi-stakeholder (patients, clinicians, academics, third sector, industry, government) buy-in, the programme has raised £150K funding to develop a national roadmap for the system. This was strongly endorsed by The Farr Institute’s International Advisory Board. A 6-month pilot starts in April and Scotland-wide implementation is planned for later in 2016. The Institute plans similar approaches to other long-term conditions and UK regions/nations.

Cancer. The Farr Institute co-hosted a Cancer Frontiers Meeting in London attended by over 100 leading researchers from the four nations of the UK, NHS, politicians, patients, and research funders in December 2015. Following this we have established a UK-wide cancer group that has combined forces to submit a collaborative bid, led by A. Renehan (Manchester) and D. Weller (Edinburgh) to the £5M CRUK Population Cancer Catalyst Fund.

Cardiovascular. The Farr Institute and BHF co-hosted a Frontiers Meeting in 2014 and in 2016 submitted a £9M proposal to the Innovative Medicine Initiative on Big Data for Better Outcomes in atrial fibrillation, heart failure and acute coronary syndromes, led by Hemingway (London), Pavis and Morris (Scotland).

Kidney. The Farr Institute established Kidney Disease@Farr in 2015 (leads Black, Aberdeen and van der Veer, Manchester). This brings together a multidisciplinary group of researchers from the four UK Farr Institute Centres together with the UK and Scottish Renal Registries, Patient View and the UK Renal Data Collaboration to generate opportunities to share ideas and learning, and foster collaborations. KidneyDisease@Farr published eight abstracts at The Farr International Conference, hosted a Frontiers meeting (Sept 2015) to scope the future research priorities, is mapping key datasets and is developing a collaborative application for programme funding that will be submitted in 2016. The ambition is to utilise The Farr Institute infrastructure to improve kidney patient care and outcomes through collaborative health informatics research.

Neuroscience. The Farr Institute research outputs on dementias as part of the Dementia Platform UK.
Enabling Data Sets and Developing New Infrastructure, Methods, Technologies and Standards for Research.

Objectives:

**Objective 2:** Enable new datasets and develop new infrastructure, methods, technologies, and standards for research described in objective 1.

**Objective 4:** Work with the owners and controllers of data to support the safe use of patient information for medical research across the UK, championing data protection, confidentiality and privacy.

**Innovative Governance**

Over the past two years, The Farr Institute approach to information governance has been endorsed by and recognised as international best practice by Scottish Government, Nuffield Council on Bioethics and the Council of Canadian Academies.

The Innovative Governance Working Group (IGWG) seeks to enable scientific research by developing proportionate governance solutions and avoiding undue burden through two inter-related strands:

1. Governance methodology research
2. Developing approximated proportionate Governance Standards for data use.


This work led to the introduction of a new unified system of Information Governance for NHS Scotland in 2016. Under this new system the CEOs of Scottish Health Boards and Caldicott Guardians have delegated decision making powers to a new Public Benefits and Privacy Panel (PBPP). This panel, launched in February 2016, comprises members of the health service, The Farr Institute research community, and lay people. A proportionate risk model is deployed involving two tiers. Tier one meets fortnightly to review all applications and tier two meets quarterly to consider more complex cases.

Lea (UCL) is working with the Information Governance Alliance and the Health and Social Care Information Centre in the development of a new anonymisation standard to complement the Information Commissioners Office (ICO) code of practice on anonymisation for health and social care organisations. He is also developing a Farr Institute-specific use case for inclusion in the standard, in collaboration with the ADRC. Lea is also the Institute’s link on governance good practice with the Alan Turing Institute.

In the Global Alliance on Genomics and Health, Laurie (Edinburgh) chaired the Task Team on Privacy and Security. The Task Team are charged with producing the relevant policy for this consortium of 200+ members involved in genomic medicine worldwide. Under his direction, and with Dove (Edinburgh and McGill) as coordinator, the international and interdisciplinary team has now produced a full policy that was signed-off at the Global Alliance annual meeting in June 2015. (Policy: genomicsandhealth.org/work-products-demonstration-projects/privacy-and-security-policy.)

Jones (Swansea) completed a data governance review to inform health informatics developments in Ontario.

Key outputs of the IGWG:

- Published ‘On moving targets and magic bullets: Can the UK lead the way with responsible data linkage for health research?’ Int J Med Inform, 2015.
- Department of Health report Protecting Health and Care Information and the consultation on The role of the National Data Guardian for health and social care.
- Council of Canadian Academies report: Accessing Health and Health-Related Data in Canada (2015) [Laurie (Edinburgh) was a member of the expert panel which produced the report].
Trusted Research Environments

The Farr Institute has led the UK in providing Trusted Research Environments (TREs) for the analysis of electronic health record and associated data. Farr investigators have worked with data controllers to develop and implement core components of ‘safe people, safe data, safe systems’ to provide the security and privacy protection that is not only required by law but is also fundamental to the honest brokerage position of The Farr Institute, working with NHS and a range of industry partners. In 2015/16 The Farr Institute established ISO27001 certified safe havens at UCL, Dundee, Edinburgh, Glasgow and Swansea. We anticipate ISO27001 accreditation in Manchester in 2016. Our vision is to create a federated approach whereby regional data science capability and platforms are interoperable with national data initiatives, because many of the datasets required for such research are not centralised in national repositories. Close links with local NHS organisations are essential to implement e-trials, precision medicine and learning health systems and to build capacity in data science at scale. Our approach to increasing the volume and population coverage of data available to researchers includes work on early replication of findings in multiple settings, the development and adoption of distributed analytical systems, e.g. DataSHIELD (www.datashield.ac.uk) and a form of federated ‘centralisation’ of data using UK Secure eResearch Platform (UKSeRP) where the individual sets of data are remotely controlled by a distributed group of data custodians. UKSeRP has been particularly helpful in supporting The Institute’s work on cohorts (see below). The Institute’s technical leads from all four nodes have also met with HSCIC and GeL on a number of occasions to inform the development of their TREs. This group also led The Farr London Frontiers Meeting to create a TRE for London. A special session was held on this theme at The Farr Institute International Conference 2015 with key opinion leaders Jon Fistein (MRC) and Martin Severs (HSCIC). The Institute’s expertise has contributed to the development of a number of TREs.

The Institute is supporting this federated approach whereby regional data science capability and platforms are interoperable with national data initiatives, because many of the datasets required for such research are not centralised in national repositories. Close links with local NHS organisations are essential to implement e-trials, precision medicine and learning health systems and to build capacity in data science at scale. Our approach to increasing the volume and population coverage of data available to researchers includes work on early replication of findings in multiple settings, the development and adoption of distributed analytical systems, e.g. DataSHIELD (www.datashield.ac.uk) and a form of federated ‘centralisation’ of data using UK Secure eResearch Platform (UKSeRP) where the individual sets of data are remotely controlled by a distributed group of data custodians. UKSeRP has been particularly helpful in supporting The Institute’s work on cohorts (see below). The Institute’s technical leads from all four nodes have also met with HSCIC and GeL on a number of occasions to inform the development of their TREs. This group also led The Farr London Frontiers Meeting to create a TRE for London. A special session was held on this theme at The Farr Institute International Conference 2015 with key opinion leaders Jon Fistein (MRC) and Martin Severs (HSCIC). The Institute’s expertise has contributed to the development of a number of TREs.

The Institute is supporting this federated approach whereby regional data science capability and platforms are interoperable with national data initiatives, because many of the datasets required for such research are not centralised in national repositories. Close links with local NHS organisations are essential to implement e-trials, precision medicine and learning health systems and to build capacity in data science at scale. Our approach to increasing the volume and population coverage of data available to researchers includes work on early replication of findings in multiple settings, the development and adoption of distributed analytical systems, e.g. DataSHIELD (www.datashield.ac.uk) and a form of federated ‘centralisation’ of data using UK Secure eResearch Platform (UKSeRP) where the individual sets of data are remotely controlled by a distributed group of data custodians. UKSeRP has been particularly helpful in supporting The Institute’s work on cohorts (see below). The Institute’s technical leads from all four nodes have also met with HSCIC and GeL on a number of occasions to inform the development of their TREs. This group also led The Farr London Frontiers Meeting to create a TRE for London. A special session was held on this theme at The Farr Institute International Conference 2015 with key opinion leaders Jon Fistein (MRC) and Martin Severs (HSCIC). The Institute’s expertise has contributed to the development of a number of TREs.

Other key outputs include:

- The Edinburgh Parallel Computing Centre (EPCC) at the University of Edinburgh; Parsons, Robertson (Edinburgh), Jefferson (Dundee), Pavis, (NSS) which launched The Farr Institute funded SAFE virtual organisation software analytical platform for user management in December 2015. This has created a national TRE for over 200 datasets curated by NHS Scotland for a population of 5 million. Currently there are more than 300 projects hosted by the platform which is shared by the ADRC.

- The Jisc Assent service (federated identity management) was implemented in 2015

- The Jisc SafeShare project to link The Farr Institute, ADRN and MRC Medical Bioinformatics Centres, is piloting an authentication, authorisation and accounting infrastructure (AAAI) to demonstrate how our access and identity management services can be used to control access to sensitive data. Controlled access will further enhance the UK’s research capacity in areas such as medical, clinical, health and the social sciences. The project will inform decisions on a potential Jisc national infrastructure for secure access to data for researchers across various academic disciplines. (www.jisc.ac.uk/rd/projects/safe-share)

- A link has been established between the Stoller Biomarker Discovery Centre and HeRC eLab

- Implementation of UKSeRP.

- Convening the three Academic Health Science Networks and Centres in London (Nov 2015) which have now agreed to develop TRE for London. The specification, implementation and governance of this will be informed by experience in Wales, Birmingham and Scotland.
Enabling New Access for Researchers to EHR Data Sets

During 2015 the Institute made considerable progress in supporting access to a number of new sources of routine data and national audits including:

- Collaborating with National Services Scotland (NSS), a special Health Board of NHS Scotland, Farr investigators have agreed the governance and infrastructural requirements to enable new nationwide datasets for research. This includes the national dispensed prescribing dataset of every community-encashed prescription in Scotland. Farr investigators deployed bespoke Natural Language Processing (Robertson, Edinburgh) to derive greater information from the scripts (e.g. dose, duration). The national picture, archiving and communications system (PACS) of over 23 million imaging studies since 2009 in all 38 Scottish hospitals is being curated for research purposes within the Advanced Compute Facility at the University of Edinburgh, on behalf of NSS (the data controller) by Farr researchers in Dundee and Edinburgh (Jefferson, Parsons, Pavis).

- Linked primary-secondary registry records. We have expanded the curation of phenotypes in CALIBER, with now over 50 projects.

- Critical Care MDS (Wales)

- Laboratory results LIMS (Wales)

- Referrals dataset (Wales)

- National Neonatal Research Database (NNRD) (UK)

- National Hip Fracture Database

- Real time intensive care data as part of the NIHR Health Informatics Collaborative. Establishment of governance framework and section 251 for the real time, five hospital collation of about 300 parameters per patient [Breatly, Singer, (UCL)]. This is widely seen as a culture changing set of achievements and the most advanced of the NIHR HIC projects.

Secure eResearch Platform (UK SeRP) and Distributed Access to Consented Cohort Studies

The Farr Institute’s UKSeRP (Lyons, Ford, Thompson (Swansea)) is providing increased distributed access across the UK to multi-modal data from multiple cohorts, supporting major funding initiatives in discovery science and public health.

Farr Investigators [Lyons, Thompson (Swansea)] have led the informatics development of the MRC’s Dementias Platform UK (DPUK) which is creating a UK powerhouse for dementia research designed to fast-track scientific understanding, treatments and the prevention of the disease. This £53 million public-private investment by the MRC, technology and pharmaceutical companies is designed to transform the best dementia research into the best treatments as quickly as possible. The Farr Institute is deeply involved in providing the supporting informatics with all four nodes being involved in bringing together multi-modal data from more than 30 cohorts to a widely distributed research community.

This is an excellent example of federated cross-UK team science in action with different universities leading on components: cohort characterisation and outcome adjudication (Edinburgh); Ethics, Legal and Social Implications (Cambridge); exposure data harmonisation (Oxford); Imaging (Oxford/UCL); sensor data integration (Manchester); genomics (Cardiff) and provision of the DPUK analysis portal on UKSeRP (Swansea). Figure 2 shows the modular structure of UKSeRP, accessible worldwide.

This collaborative infrastructure is seen as world leading in this field and a number of European and North American cohorts have expressed interest in joining the DPUK initiative.

UK Biobank. UKSeRP will shortly provide access to multi-modal linked UK Biobank data to support the working of the distributed outcome adjudication team.

Millennium Cohort Study (MCS). UKSeRP is being used to provide access to linked data from MCS sweeps 1-5, physical activity (accelerometer) and oral fluids data, hospital admission data from England, Wales and Scotland for a Wellcome Trust funded study involving the Swansea and London nodes of The Farr Institute “Using health record linkage in the UK MCS to investigate childhood obesity, asthma, infections and immunisations in early life” (Lyons, Dezateux).

Alcohol Misuse Electronic Longitudinal Alcohol Study in Communities (ELASTIC) is an ESRC funded project to explore the impact of varying levels of alcohol intake across the life course utilising data from a number of traditional and anonymised population based e-cohorts, including UK Biobank, Understanding Society/ British Household Panel, ALSPAC, MCS, e-CATALYST and the Wales Electronic Cohort for Children (WECC).

Linkage to Consented Cohorts to EHR Sources

Progress with cohort study linkage has been quite mixed during 2015 despite considerable development and progress with bringing together and providing remote access to baseline and study enhancement data through Farr’s UKSeRP. Cohort linkage to routine data has been much slower but is now starting to show progress following reorganisation of functions and the bedding down of new processes at HSCIC. A number of meetings were held between Farr Institute and those working on cohort linkage, including with the ESRC funded CLOSER initiative (an online resource that enables researchers to view and appraise data from eight leading UK longitudinal studies), the Wellcome Trust brokered meeting with the HSCIC and separate meetings in Edinburgh, Cambridge and London on engagement with cohort custodians as part of the DPUK initiative.
Methods & Tools

Methods in record linkage. There have been developments in probabilistic record linkage tools (Enhanced Matching System) which without the use of an NHS number demonstrate excellent sensitivity (99.5%) and specificity (100.0%). Work on Bloom Filter matching (matching of hashed identities) with Curtin University (Australia) is also underway.

There is also considerable activity on going on natural language processing (NLP) of text data in a number of nodes, the implementation of common data models, harmonization and sharing of variable definitions and concepts. Farr Institute has adopted www.clinicalcodes.org to provide a repository for clinical code lists for phenotype definition as part of its infrastructure. Farr Institute has also recently initiated multiple projects using machine learning for different aspects of EHR research such as disease phenotyping, risk prediction and unsupervised classification approaches for pathway identification and extraction. ClusterExplorer is an opensource tool written in R and developed by Farr Institute for creating interactive dot maps which was designed for use by public health practitioners and researchers to assist with disease cluster investigations.

Analytics capacity in partnership with the Alan Turing Institute (funded by EPSRC), the MRC Medical Bioinformatics Initiative and with further support from EPSRC, The Farr Institute has established the UK Health Data Analytics Network (UK-HDAN) led by Taylor (Manchester) and Hogg (Leeds). The Network is a broad community of health data scientists working together to map the UK health data analytics landscape and to frame a national strategy for health data analytics research. In January 2016 an inaugural workshop was held in Manchester and follow-up events have been scheduled in Sheffield and Birmingham. The Network will focus on data science to address key challenges in health and social care, bridging the gap between the development of foundational theory, methodologies and algorithms, and the use of electronic health records and other population-based data for health research and service improvement.

In collaboration with the MRC Medical Bioinformatics award holders The Farr Institute supported eMedLab, a collaborative data science environment that brings together UCL, Kings, LSHTM, The Crick Institute, Sanger and EBI/EMBL (lead Luscombe (UCL, Crick)). This shared computer cluster integrates and shares heterogeneous data from personal healthcare records, imaging, pharmacoinformatics and genomics. It also provides a platform to develop analytical tools that allow biomedical researchers to transform raw data into scientific insights and clinical outcomes whilst storing data securely. Its modular design will ensure sustainability through expansion and replacement.

Farr Institute High Performance Computing facilities at UCL are being used by the ICONIC project (£3.6M Wellcome Trust/Department of Health Innovation Challenge Fund award) for viral genome assembly and analysis.

The eLab (eInfrastructure theme, HeRC@Farr) is now established as a cornerstone of a wide range of Farr researcher-led bids, for example the £5m CRUK catalyst led by Renehan (Manchester) in collaboration with The Farr in Scotland and successful grants such as Childhood Wheeze (Wellcome Trust £4.5m led by Custovic, Manchester), WHEAT (MRC £800k led by Gale, Imperial). HeRC (Custovic) and Imperial are leading work in a £10m Euro H2020 bid (UNICORN) to optimise population cohorts that will see The Farr Institute eLab integrated with at least two other platforms holding asthma cohort data to maximise analytical power. HeRC will lead the distributed analysis (Buchan, Couch).
Objective 3: Develop skills, talent and expertise in individuals and foster research communities for collaborative working.

The Farr Institute has transformed the opportunities for early, intermediate and senior researchers and practitioners to gain new skills in data science. Currently, there is a skills shortage across a number of areas of expertise where we need skilled individuals to work together to solve the challenges of obtaining meaningful and accurate information from heterogeneous data types. The Capacity Building Working Group [McCowan (Glasgow), Taylor (UCL), Moulton (Manchester), Anastasiou (Swansea), Goddard (Dundee) supported by Paul Stewart (Leeds)] lead on the implementation of The Farr Institute's strategy to build a new breed of 'health data scientist' with an understanding of both big data and health care. In 2015/16 we have achieved this through:

The Farr Institute doctoral training programme (DTP).
There are 78 graduate students engaged in PhDs related to health informatics research across the Institute’s Centres. The first Farr Institute PhD symposium was hosted in June 2015 when 26 students gathered in Manchester. The aim was to introduce students to their peers across The Farr Institute and begin to develop relationships which we hope will extend throughout their careers. Each student presented their research project through an oral or poster presentation. Students also learned media communication skills and recorded a 2-minute video summarising their research and future aspirations. These videos now form a collection which can be viewed by the public through The Farr Institute’s YouTube channel (goo.gl/G173LX ). The PhD Symposium and Summer School will take place in 2016.

Masters programmes. Taught Post-graduate and Online Distance Learning opportunities have been developed at Glasgow, Edinburgh, Manchester, Swansea and UCL across a range of health informatics and data science disciplines. Modules include: Practical Introduction to High Performance Computing; Social Shaping of Digital Research; Engaging with Digital Research; The Use and Evolution of Digital Data Analysis and Collection Tools; Technologies of Civic Participation; Understanding Data Visualisation; Managing Digital Influence. These MSc programmes in Health Informatics and Data Science have recruited 38 (Swansea), 12 (UCL), 9 (Manchester) and 10 (Edinburgh) new students in 2015 bringing total enrolled since course commenced to 76.

Short courses and masterclasses to allow professionals to gain knowledge in new fields. 350 people attended Farr Short Courses and CPD in 2015. Participants came from a wide range of professional roles and back grounds including clinical lecturers, clinical scientists, professors, research fellows, PhD students, NHS consultants, NHS registrars, information officers, project managers, data analysts, data mangers, statisticians, software developers and programmers and system developers.

Worked with the NHS to build informatics know how across the workforce; for example, in Scotland the Chief Statistician of the Scottish Government has subscribed to Farr Institute courses for Government analysts and funded ten exemplar projects to demonstrate the impact of cross sectoral data linkage in collaboration with The Farr Institute and ADRC. Manchester and UCL have led in the curriculum development and delivery of NHS-focussed clinical scientist programmes in Health Informatics at MSc (Scientific Training Programme) and PhD (Higher Scientific Specialist Training) levels. Five students are currently enrolled on the MSc STP.

Contributed to the Academy of Medical Sciences report ‘Improving recognition of team science contributions in biomedical research careers’ (March 2016) which recognised that the current academic career path is a hurdle to those with a specialised skill (often methodological or technical) who fall outside the normal Principle Investigator role but are required for the successful delivery of interdisciplinary collaboration.
Objective 5: Engage with the public to demonstrate the benefits of using health data in research and to encourage the support of secure and trusted access to patient information.

Patient and public involvement and engagement (PPIE) is central to securing public support for research using health data and improving the quality, impact and relevance of our research. With PPIE embedded within each Centre, we are now focusing on developing an Institute-wide strategy and delivering gold standard PPIE that enriches, supports and enables ground-breaking research, in line with The Farr Institute’s objectives.

Support for #datasaveslives campaign.
The Farr Institute has launched #datasaveslives (#dsl), a public engagement campaign designed to highlight the positive impact of health informatics research on public health. Generating over 1,700 mentions to date, the campaign has gained support from Ministers and politicians (George Freeman MP, Shona Robison MSP, John Barron MP), the public, policy leaders (Bedirhan Ustun, WHO) and researchers alike. The campaign has been picked up by other organisation such as CRUK, Public Health England and the government backed #empower campaign for health data use in research. #dsl has also been adopted by the £20 million Department of Health funded Health North pilot Connected Health Cities as their leading tool for public engagement.

Public panels are well-established in all four Centres, constituting a unique network of over 50 regular public contributors, well-informed in health informatics and data linkage research methods. Panel members review and collaborate on research proposals from an early stage, some as co-applicants, strengthening involvement from conception to delivery. Public input has shaped proposals to funders including The Wellcome Trust, CRUK (CREATOR, >£5m) and NIHR (pragmatic trials, van Staa).

CIPHER host the UK MS Register which aims to address the lack of reliable evidence-based information on MS so that it can be used for further research and to inform policy and practice to help people with this condition. The UK MS Register team exhibited at the MS Life conference, attended by over 3,000 people with MS and their families. The Register is a collaborative entity, therefore we have created a ‘People Affected by MS’ panel which is crucial to the Registers research and focus.

Public input shapes ongoing research, often negotiating sensitive topics.

For the Dementia Platform UK project, over 30 patients, carers and members of the public tested wearable activity trackers and discussed sharing data for dementia research at interactive workshops. Findings directly informed purchasing decisions for wearables to support future projects.

Farr London have developed a public engagement programme among hard to reach populations including migrants and homeless populations. A recent public engagement workshop ‘What works in inclusion health: evidence synthesis and engagement on effective interventions’ involved people with lived experience of social exclusion and marginalisation as co-researchers in the interpretation and writing up of a Lancet Series.
We are using innovative participatory research methods to involve wider sections of the public, including citizens’ juries and large-scale citizen science (e.g. Cloudy with a Chance of Pain), to explore public opinion, harness trust and engage the public in the ‘doing’ of research.

Collaborating with capacity building colleagues, PPIE and research staff have delivered sessions in schools, including an inspiring day for 40 school children at the 2015 Farr conference and at Manchester’s prestigious Altrincham Grammar school.

We have delivered interactive engagement activities at major public events, including: Edinburgh Science Festival, Edinburgh Skeptics Society, UCL Digital Science Festival, March 2016 (Farr London workshop for researchers and lay members of the public to deepen understanding of the reasons for, and impact of, missing information in the electronic health record from patient and researcher perspectives) and Manchester Science Festival (>1000 people visited interactive platforms on data sharing).

Public Opinion Research Informs Policy: Health Data On Trial Using Citizens’ Juries

In January 2016, Manchester hosted two ‘Citizens’ Juries’. ‘Juries’ of lay members heard expert opinion over three days, tasked with answering the question “to what extent should patients be in charge of their own records?”

On the 10th March, Farr and the ICO will co-host the launch of the report to an audience of the UK’s top policymakers, academics and Dame Fiona Caldicott (National Data Guardian) and discuss policy implications for sharing data for public benefit. A summary has been commissioned to inform the forthcoming National Data Guardian Review.

Born In Bradford (BiB)

HeRC provided BiB with 100 phones funded by The Farr Institute for research to maximise impact in their community. Children from a local school were shown how to use the phones as cameras which they then used to photograph the food they ate. Poet Ian McMillan worked with the children to compose a ‘favourite food’ poem using this data which they performed with the two Ians at the BiB Scientific Conference in Sept 2015.

In an annual programme BBC Radio4 make about BiB, the phones were used as recorders, getting the children to do their own interviews with family and friends, resulting in a delightful 30-minute documentary charting their lives, languages, learning and loves in Bradford. www.borninbradford.nhs.uk/the-bib-filmphotography-and-poetry-gallery/radio-gallery/

Dissemination takes place through a variety of channels: popular birthday cards, Facebook, newsletters (to 12,500 homes) children’s centres, GPs, schools, summaries of papers on the website and national media. BiB has a parent governors group who have become expert in providing insights into research proposals and dissemination; many of the HeRC@Farr related papers have benefited from their input and local knowledge.
Objective 6: Bring together government, public sector, academia and industry to foster relationships and establish best practices for innovation, discovery and impact in health data science.

Disseminating New Research, Methodology & Innovation
The Farr Institute hosted two major international conferences in 2015.

The Digital Health Assembly Open Innovation Conference, Cardiff, 10-12 February.
Conference themes:
- Big Data
- Empowering Patients and Staff
- Innovative Business Model

Attended by 359 thought leaders and practitioners in the fields of digital health and open innovation including industry, NHS, government and academia from the UK and overseas (USA, Europe and Asia).

A mix of keynotes, roundtables, panel discussions and interactive workshops provided ample opportunity for knowledge exchange and networking to generate innovation.

The Farr Institute International Conference 2015: Data Intensive Health Research and Care, St Andrews 26-28 August.

Attended by 450 delegates from Australia & NZ (24), Canada (33), Europe (6), USA (6) and UK (364) including 70 PhD students.

Health practitioners, data science researchers, policy makers and members of the public were all represented at the conference which hosted 275 Oral and poster presentations.

Invited speakers:
- Dr Charles Friedman, University of Michigan
- Anne McKenzie, The University of Western Australia School of Population Health
- Dr John Mattison, Kaiser Permanente, SCAL
- Dr Michael Schull, Institute for Clinical Evaluative Sciences, Canada
- Professor Ruth Gilbert, ADRC–England
- Prof Sir Munir Pirmohamed, University of Liverpool
- Prof Sarah Garner, National Institute for Health and Care Excellence
- Prof Cathie Sudlow, UK Biobank
- Prof Martin Severs, Health and Social Care Information Centre
- Prof Sir Alex Markham, MRC Medical Bioinformatics Centre, University of Leeds
- Prof Yike Guo, Data Science Institute, Imperial College London

In 2016 The Farr Institute is joining with the International Population Data Linkage Network (IPDLN) and Administrative Data Research Network (ADRN) to co-host the IPDLN Conference which will be held in Swansea and is Chaired by Ford (Swansea) the current Director of the IPDLN. Themes for the conference include: Analytical approaches to distributed data; Data and linkage quality; Advanced analytics, linking to emerging data types; Privacy, regulation and governance; Public engagement; Capacity Building.

Participation in other meetings
EU-wide level trials methodology dissemination undertaken at scale at the ‘GetReal ’consortium annual meeting held in Manchester over 2 days in February 2016 (around 100 delegates), with a strategic pre-conference meeting with National Institute for Health and Care Excellence (NICE), both managed and coordinated by The Farr Institute Operations team based in Manchester.

The Institute was invited to host and co-organise The ‘Machine Learning for Personalised Medicine’ (Lawrence, Sheffield) summer school. Held at Manchester’s Museum of Science and Industry this saw 5 days of seminars, workshops, posters, pc-based workshops, talks and networking for around 80 delegates from across the domain.

Manchester is European City of Science in 2016 and The Farr Institute’s involvement has been proactively sought by organisers, including two projects (Cloudy with a chance of pain & #britainbreathing); an industry-academic panel session “Trust me I am data” (Hassan, Buchan, van Staa, Shiff, MSD) and a range of PPIE events during the festival in July 2016.

An Institute-wide panel session was accepted for the Royal Statistical Society’s 2016 conference (Manchester, Oct 2017) on the application of statistics in health research (Sperrin; Berridge; Williamson; Kontonpantelis; Peek; Buchan)

Following on from The Farr Institute Cancer Frontiers Meeting, London 2015, the Institute was invited to co-host a follow up meeting with NCRAS and invited to host a session at the NCRI Conference, November 2016.

The Farr Institute has been invited to give an opening keynote and a number of additional talks at the BioData World Congress 2016 (Hinxton, England).
Raising The Farr Institute’s Profile Through Social Media

www.farrinstitute.org underwent a rigorous redevelopment with a new site launched on 13 April 2016. Key improvements included:

- Revised ‘About’ section including ‘Who We Are’, ‘What We Do’, ‘Our Funders’ and ‘International Advisory Board’ pages
- Research and Education
- Searchable database of ‘our investigators’
- Pages and subpages to highlight key research themes
- ‘Recent Publications’ page in News section
- Public Engagement section
- ‘100 Ways of Using Data to Make Lives Better’ page to highlight case studies
- News and Events
- Enhanced news, event and courses listings
- ‘Press and Media’ page to encourage media enquiries and speaker invitations
- ‘Features and Newsletters’ page to showcase articles, opinion pieces, interviews and guest blog posts
- ‘Working with Us’ section to encourage collaboration enquiries

The Farr Institute has actively embraced social media as a means of communicating with the public and professionals. The Farr Institute and each of its regional Centres have twitter accounts and have attracted over 4,000 followers. The Farr Institute’s LinkedIn discussion group more than 200 followers. The Institute is developing a broad portfolio of public facing online videos which include:

- Schools Engagement Workshop, International Conference 2015 www.youtube.com/watch?v=OeoEoOvP7rQ
- #datasaveslives www.youtube.com/watch?v=jWfuPTmUti4
- Keynote Speakers from the International Conference 2015 www.youtube.com/watch?v=hYu6CvRkBA
- PhD Interviews, PhD Symposium 2015 www.youtube.com/watch?v=LVlPYGw33lw
- A public-facing YouTube video was also created of PhD students at the Born in Bradford (Bib) conference in September 2015. www.youtube.com/user/BorninBradford2015/videos

The Farr Institute in the Media

A number of Farr Institute-supported outputs were successfully driven to wider national and international audiences through media coverage. These include the HeRC@Farr collaboration with Withings on the relationship between weight management and the use of smart scales (Sperrin) which was picked up by Le Monde & Men’s Fitness www.mensfitness.com/life/gearandtech/5-best-smart-scales

The HeRC@Farr hub also supported the launch of the Cloudy with a chance of pain project (Dixon, ARUK). Channels included national television and press coverage including the Mail Online and The Guardian; TV coverage on BBC breakfast and Trust me I’m a Doctor and ITV news, filmed at The Farr Institute’s Vaughan House location.

In February 2016 Director Prof Harry Hemingway was featured in a BBC News Health article ‘The Challenge of Saving Lives with ‘Big Data’; published 7 Feb 2016 www.bbc.co.uk/news/health-35491177

In March 2016 Director Prof Andrew Morris featured in a 30 minute BBC Radio Scotland Brainwaves programme dedicated to The Farr Institute, broadcast 15 March 2016 www.bbc.co.uk/programmes/b0745jd4

In April 2016 The Guardian newspaper and HSCIC hosted a debate: Can Technology Save the NHS? Featuring comments from Prof Harry Hemingway


Dr Elizabeth Williamson (LSHTM) discussed the statistics of National Lottery wins on PM with Eddie Mair, BBC Radio 4, broadcast 7 Jan 2016 www.bbc.co.uk/programmes/b06fwhfq
Partnerships and Collaborations

Farr Institute Objective 6: Bring together government, public sector, academia and industry to foster relationships and establish best practices for innovation, discovery and impact in health data science.

The creation of an outstanding health data science ecosystem ultimately dependent on building trust and joint strategy with a broad range of NHS, Government, industry and third sector organisations. As The Farr Institute is not a data controller, it is vital that time and resource is dedicated to establishing strong working relationships with a variety of bodies across the UK. The Farr Institute has hosted high level day-long delegations with a variety of UK leaders including, from NHS England (Sir Malcolm Grant, Chairman), Innovate UK (Ruth McKearman CEO), HSCIC (Andy Williams CEO), Office for Life Sciences (George Freeman MP, Minister for Life Sciences) and the Scottish Council for Economic Advisors (Nicola Sturgeon, First Minister, and John Swinney, Deputy First Minister). Here we report on key activities in 2015-2016.

NHS Partnerships

Public Health England. e.g. on global burden of disease work and Cancer informatics programme (David Forman and Jem Rashbash).

HSCIC. e.g. Daniel Ray (Farr London) was appointed Director of Data Science for HSCIC in March 2016.

NHS Scotland. Led by Dr Aileen Keel, former Deputy Chief Medical Officer, the Scottish Government has established an Innovative Healthcare Delivery Programme (IHDP) within The Farr Institute to fundamentally change the way data can be used to drive improvement in health outcomes by fostering new relationships and connecting the NHS, academia and industry. It is supported by up to £3M from the Scottish Government.

National Services Scotland. Twenty NHS staff from the Public Health Intelligence staff of NSS are co-located within The Farr Institute in Scotland.

NIHR Biomedical Research Centres in England renewal (UCL has 3: UCLH, GoSH and Moorfields), new applications Barts, QMUL, Manchester, Liverpool. Each of these has Data science/informatics themes (led by Farr Directors). Collaboration across England in the development of shortlisted full proposals will be a significant catalyst particularly for hospital informatics.

The Farr Institute funded rollout of Northwest eHealth clinical trials tool FARSITE continues at pace. This partnership with stake-holders across the North’s Academic Health Science Networks (ASHN) is a direct result of investment from The Farr Institute capital spend (£500k) enabling clinical research networks and AHSN teams to utilise the FARSITE platform for service improvement and efficient trial recruitment.

The Prudent Healthcare Intelligence Unit is being established at Swansea University. This Unit will focus on working directly with NHS staff and analysts to carry out service-focused research to for NHS service improvement. The co-location of NHS analysts in the Data Science Building will enhance their interactions with the SAIL database, thereby increasing the NHS’ capacity in data linkage research and the ability to measure patient and population outcomes.

Partnerships have also been established or strengthened with the Public Health Wales Observatory, the Welsh Cancer Intelligence & Surveillance Unit, the All-Wales Therapeutics & Toxicology Centre, a number of local authorities and Welsh Government department.

Industry

15 projects are currently in development ranging from student placement consultancy, through to multi-site observational studies and new technology solution development and optimization. Many large organisations are also proactively seeking out to engage with the Institutes, and such approaches have been made by companies including Hewlett Packard, IBM, Intel, GE, Astra Zeneca and Ipsen Pharmaceuticals.

Astra Zeneca has funded an inter-related series of projects (APOLLO programme) relating to the real world evidence for prolonged dual anti-platelet therapy. This includes: 4 country international comparisons of care and outcomes (US, UK, Sweden and France in 140k patients using EHR, Rapsomaniki et al 2016), mapping the key trial inclusion and exclusions criteria to unselected patients in English EHR (Timmis under review 2016) and developing and validating prognostic risk models for both benefits and harms (bleeds) and demonstrating for each individual net clinical benefit (Pasea ACC 2015).

The head of Astra Zeneca’s Advanced Analytics Centre, Dr Jim Weatherall has joined Farr as a visiting Farr researcher based in Manchester. This will advance collaboration and co-creation in the area of real-world medication optimisation, specifically in capacity building, trials methodology and endotype discovery and is our first example of a spin-in collaboration.

Manchester hosted a two-day workshop with US and UK representatives from Merck Sharp & Dohme (May 2015) with around 25
attendees. MSD’s Realworld Evidence and Observations Team (led by Susan Shiff) and the Farr HeRC team explored synergies, priorities and laid solid groundwork for future collaborations. Tangible outcomes as a result of this relationship include Buchan being invited to speak at MSD in New Jersey in April 2016.

The Farr Institute in Swansea is working with a number of companies to use linked data to describe, evaluate and predict health and healthcare trends for health boards and to develop software that allows clinicians and NHS managers to ask complex questions about health services using familiar terminology without needing an indepth understanding of the data and systems.

The Farr Institute has built upon the existing strategic relationship with Intel over the course of the year, with the company contributing manpower and equipment to projects running in Farr HeRC, Farr London and Farr Scotland. Intel assisted in the facilitation of a Farr hosted work-shop in Cardiovascular Imaging with more than 30 academics and companies participating, and a consortium has formed with a view to developing some of the proposals put forward at the workshop.


Our collaborative working with ABPI in 2015 focused on roundtable events, and in 2016 we are looking to under-take some more thematic joint workshops with quantifiable outcomes.

The Farr Institute (Buchan, Custovic), through SME-Academic partnership and mentoring, has been recognised by the Biorelate CEO (Daniel Jamieson) in being instrumental in their successful securing of key funding in Feb 2016. The Farr HeRC relationship with local SME Melandra has also evidenced the impact SME expertise on large scale capital projects (CPC, Ainsworth).

Towards an International Strategy

The Farr Institute is committed to contribute to working together with international partners to create a common framework of harmonized approaches to enable the responsible, voluntary, and secure sharing of health related data, and to form creative and high energy international research collaborations. We have started to work systematically with key opinion leading organisations including:

- Global Alliance for Genomic Health. Morris represents The Farr Institute on GA4GH, is member of the Clinical Working Group and co-chairs the eHealth subgroup. The Clinical Work group aims to enable compatible, readily accessible, and scalable approaches for sharing clinical data and linking it with genomic data. Recognizing how much work is already ongoing in this area, the Clinical Working Group seeks not to reinvent the wheel, but rather to add value to existing endeavours. Work products include an international catalogue of major eHealth research programmes, a family history inventory and family history toolset.

- In 2015 The Farr Institute has agreed a strategic relationship with the Institute for Clinical Evaluative Sciences, Ontario, Canada formalised by MOU. It is expected that this partnership will lead to collaborative capacity building and research as well as sharing best practice in public engagement and methodology workstreams.

- In 2015 Farr has strengthened relationships with the University of New South Wales (UNSW) in Australia and Harvard Medical School. We have delivered reciprocal masterclasses and development activities in addition to research collaboration.

The Farr Frontiers Meetings

The Farr Institute convened 11 collaborative meetings during the year focusing on a number of research groups and clinical topics, including: asthma, renal, cardiovascular, cancer, primary care, drug safety, dental health and mental health. These meetings welcomed academics, clinicians, industry and third sector partners from beyond The Farr, with a view to broadening the reach of the Institute. The meetings have catalysed the creation new collaborations and four further funding bids. For example, the Cancer Frontiers meeting led to the formation of a consortium led by Farr Manchester for the CRUK Catalyst award. York’s haematological malignancies research team, and Manchester’s Biostatistics group are also building a collaborative proposal as a result of the event.

Manchester-Harvard Exchange Programme

The first of its kind, the workshop, hosted at Farr HeRC’s Vaughan House, combined collaboration with career development by enhancing independent research skills and providing the opportunity for early stage researchers to develop their own novel health informatics ideas. Funded by a Manchester based partner (MIMIT), HeRC was able to extend funding to allow a Farr researcher to join the process (Jones, Liverpool). Through an innovative hands-on model of idea generation and free thinking, attendees identified new ideas organically. A formal competition is now underway to fund the strongest idea, with a trip to Harvard to shape a funding proposal in collaboration with Harvard colleagues.

www.herc.ac.uk/case_studies/harvard-medical-school-partners-herc/
2015 has been an exceptionally busy and successful year for The Farr CIPHER node. There were 105 publications, 29 grants secured (value £19.9M) and 123 presentations and engagement activities. A number of Farr CIPHER-led activities are reported elsewhere in this report.

New Investment & Capacity Building
Farr CIPHER and DECIPHer (MRC Public Health Improvement Research Centre) investigators awarded one of six Health and Care Research Wales funded centres, the National Centre for Population Health and Wellbeing Research, which will add to and complement the activities of both centres. Additional data analysts have been funded through formal collaborations with the Wales Cancer Research Centre, Wales Kidney Research Unit, National Centre for Mental Health and the Wales Centre for Primary and Emergency Care Research. 2015/16 saw a significant increase in the number of students on the Swansea-based Masters courses in Health Data Science (from four students in 2014/15 to 17 in 2015/16 [325%]) and Health Informatics (from 18 to 28 students [56%])

Policy Influence
A range of research projects designed to influence policy have been carried out and results discussed with policy makers, including a study which found positive effects of the Welsh Government’s Primary School Free Breakfast Initiative, and a Lancet paper analysing whether mortality rates are higher at weekends in England and Wales across fifteen different conditions. Communications and engagement activities with Welsh Government departments is being increased substantially. For example, a joint CIPHER/ADRC-Wales engagement event for policy-makers from across Welsh Government is being scheduled. The number of embedded policy research officers is also expected to increase.

Communications to Enhance Research Impact
Farr CIPHER has been working to increase capacity and effectiveness in communicating research findings and stimulating further research.

- A Communications Officer has been recruited to create a significant number of case studies as well as brochures targeted at specific users of our research (e.g. NHS Health Boards).
- An NHS Engagement Officer has been recruited to build more and deeper relationships with NHS organisations.
- Investment in training NHS analysts and engagement activities has led to NHS organisations contributing data analysts to the Prudent Healthcare Intelligence Unit and liberating more datasets for joint service improvement and research activities.

In 2016 Farr CIPHER and ADRC-Wales jointly hosted a visit from officials from the UK Department of Business, Innovation and Skills to follow up on the success of investments through the MRC and ESRC and how these support policy relevant research and produce economic impact.

Public Involvement and Engagement
Farr CIPHER has established a strong ethos of Patient and Public Involvement and Engagement and has a citizens/consumer panel which engages across Farr CIPHER, ADRC-Wales, and the SAIL Databank. In 2016 we intend to establish a pan-Wales Public Engagement group comprising all major stakeholders of Farr CIPHER, ADRC-Wales and SAIL (e.g. Welsh Government, third sector organisations) to ensure a strategic, coordinated and high coverage level of public engagement across Wales.

Cross-Farr and Wider Collaborations
2015 also brought a marked change in the breadth and scale of cross-centre activities and wider collaborations. The following is a summary of the highlights and future aspirations, using the headings devised by The Farr Institute and funders in 'A New Approach: National Health and Bio-medical Informatics Research Institute', reflecting our commitment to this aspiration.

Discovery Science
We have initiated a collaboration with UCL (Farr London and eMedLab) on immediate replication of findings on cardiovascular disease mechanisms combining data from CALIBER, SAIL and CPRD and utilising traditional statistics and machine learning approaches. The Cardiff-based CREATE Centre (Arthritis) is also part of the Manchester led MATURE consortium, a new MRC/ARUK Stratified Medicine
funded group identifying therapeutic biomarkers. Farr CIPHER provides the remotely accessible data analytics platform to integrate multimodal data from 30+ cohorts, supporting discovery science and the identifications of biomarkers for the MRC’s Dementias Platform UK, utilizing the capital investment from The Farr Institute and IDEA (Integrated DementiaA research environment) awards.

**Precision Medicine**

Notable developments include combining genomic and health record data in the PsyCymru hybrid psychosis cohort (with MRC Centre for neuropsychiatric genetics and genomics) and the work on ICONIC (Improving Cancer Outcomes in the NHS through Integrated Computing) that will combine cancer registry, NHS clinical data, SAIL data sets with genomic analyses undertaken by NantHealth (NantHealth.com).

**Public Health**

We continue to develop a series of large scale population based e-cohorts covering different age groups, diseases and disorders, and have started to develop matched e-cohorts with colleagues in Scotland (learning difficulties, autism) and London (child pharmaco epidemiology, combining SAIL and CPRD data). CIPHER investigators lead Healthwise Wales, a new general population cohort seeking to recruit 250,000 individuals. Work on NIHR funded evaluations of natural experiments (Housing and Health; Alcohol) and others (Air Pollution alerts) are nearing completion, some with clear counterintuitive results that challenge current orthodoxy. We also contribute substantially to the Global Burden of Diseases project. A number of school based physical activity and educational improvement interventions are underway supported by the MRC DECIPHER Public Health Centre, ESRC and other funders.

**Citizen Driven Health**

This area is still quite developmental. Our view is that most of the technology is still too immature for widespread adoption and implementation but we expect this to change rapidly and will seek to partner with promising developments, particularly through our DPUK links with Farr Manchester, who lead on sensor development and integration.

**Learning Health Systems**

This area has developed extensively, particularly through growing partnership with NHS policy and practitioner communities and collaboration with wider research groups, including The Farr Institute’s renal group (all four centres), Asthma UK Centre for Applied Research, and diverse specialties including critical care, emergency care, orthogeriatrics, oncology, neurology, respiratory, gastroenterology, cardiovascular, metabolic disorders, paediatrics, mental health and primary care groups. CIPHER investigators led a panel discussion on e-trials at the Health Technology Assessment International conference in Oslo. The results of the e-enhanced SAFER2 and PRISOMATIC trials will soon be released and new evaluations have been funded, including TIER (Transient Ischaemic Attack 999 Referrals), ESPRIT (enhanced service evaluation) and ERA (Electronic Records in Ambulances).

**Leadership**

Farr CIPHER investigators contributed to UK and international debates on improving the quality of source clinical records, privacy protection and the use of linked records for public good, including evidence provided to the House of Commons Science and Technology Select Committee Inquiry on Big Data, expert reviews for the ESRC Research Ethics Framework, involvement in multiple frontiers meetings, and at international consensus meetings.

**Skilled People**

Farr CIPHER staff contributed substantially to The Farr Institute PhD Symposium in Manchester and summer school in St Andrews, helping to design, develop and deliver the health visualisation components. ESRC has funded a number of ADRC studentships linked with Farr investigators. In addition, Farr CIPHER is linked to the newly awarded EPSRC Digital Economy Centre in Swansea – CHERISH-DE (Challenging Human Environments and Research Impact for a Sustainable and Healthy Digital Economy), focussed on bringing early ideas to fruition swiftly though prototype and pilots, and supporting cross-disciplinary capacity building opportunities.

**Core Methods**

We have continued to develop methods and platforms to support the development of trusted research environments, data sharing and analysis. Both SAIL and The Farr Institute’s UK Secure eResearch Platform (UKseRP) received ISO27001 accreditation in 2015. The Sussex/Brighton/Swansea NLP group have been particularly active, producing a systematic review of the utility of free text in case detection and the Harvey Corpus, the first set of syntactically annotated primary care records. The group utilises IBM Watson and Clinithink’s new Clix software on anonymizing and redacting identities and creating meaningful deep phenotyping data and are exploring the potential use of this software in the efficient identification of trial participants. A collaborative grant with CIPHER investigators at Curtin University (Australia) on development of probabilistic matching of hashed identifiers has been submitted.

**Partnerships**

The number of organisations with which we have meaningful partnership arrangements continues to grow, particularly around the NHS Prudent Healthcare and care pathway optimisation agendas. Aneurin Bevan University Health Board and Cwm Taf University Health Board now have embedded analysts working with us on the evaluation of their Living Well-Living Longer (CVD secondary prevention), Bump Start (prevention of obesity in pregnancy) and maternal smoking prevention initiatives.
Farr Health eResearch Centre (HeRC), Led by University of Manchester.

The Farr Institute HeRC has five research themes and four enrichment themes (reported in other sections).

Co-Produced Health

Human-computer interaction methodology for managed self-care has been presented in leading conferences and journals and applied to mental health, musculoskeletal, renal, cardiovascular, and public health (physical activity).

Cloudy with a Chance of Pain (Dixon, ARUK) has recruited over 8,200 UK participants to study the relationship between weather and musculoskeletal pain using smartphones. HeRC has run the project management, communications and public methodology. The project has had substantial national press and TV coverage since its launch from in February. www.cloudywithachanceofpain.com

City Verve. HeRC leads the health theme of the UK’s Internet of Things £10m demonstrator (Innovate UK, Taylor, van Staa, Peek) for Greater Manchester Combined Authority with 21 partners spanning NHS, social care and industry (e.g. Cisco) to demonstrate at scale, the use of different communication protocols, sensors, platforms, data hubs and applications to connect data, citizens and behaviours more intensively in a city region, aiming to improve services and experiences.

Healthy Balance Group. HeRC established a Greater Manchester-wide group of stakeholder organisations committed to reducing obesity. This group has been invited by British Heart Foundation to study novel social technologies for helping children achieve a healthy weight (Peek, Greenstein). ITV news reported on this through BHF-funded researcher, Adam Greenstein: www.youtube.com/watch?v=9osVbxKjpP8

ClinTouch, www.clintouch.com, mobile psychometric technology, developed for people with serious mental illness, was spun out as Community Interest Company in Dec 2015. ClinTouch was a finalist in the 2015 BioNow Healthcare Innovation Awards and won the Outstanding Innovation Award 2015 Unlimited/SEE Change.

Endotype Discovery

HeRC’s blended machine learning and biostatistical methodology research has now resulted in six major findings in asthma and allergies, and the approach is being extended to stratified medicines projects in psoriasis (Peek, Buchan) and rheumatoid arthritis (van Staa, Peek, Dixon, Humphries, Hyrich).

The STELAR (MRC, Custovic, Buchan) Asthma eLab is now live and supporting multi-site, multi-dataset/cohort research, delivering ‘team science’. Methodological extensions to STELAR funded by the Institute are enabling it to be reused beyond asthma and with HeRC’s Research Object design (also taken up in the US BD2K initiative). See the ‘Enabling data’ section for more information on international interest in taking the eLab approach to research across cohorts, datasets and universities.

Health eRecord Discovery Science

HeRC-led statistical and informatics methodology for harnessing routine healthcare data, illustrated for EHR data analytics, has been reported in all of the top international health informatics conferences and journals in 2015/16. EPSRC have awarded a new clinical text mining network to a HeRC-led consortium (Nenadic) including investigators from across the Institute working on human and veterinary healthcare records to make the free text more research-ready.

Born in Bradford and the Sanger Institute are linking sequence data and genetic variants with NHS records. They have shown that more than 1,000 rare variants are not associated with additional health service use, suggesting that these are healthy individuals. The team are now following up specific variants to assess health impact. The related Better Start Bradford project (Pickett, Wright, Lottery Funding, £49m) began in March 2015, recruiting researchers and recruiting to the new BIBBS birth cohort. As the world’s first experimental birth cohort, this is charged with the evaluation of the effect of the 22 Better Start Bradford interventions on the health and wellbeing of families. Cohort follow-up is almost entirely through linkage to routine data, including health and education www.borninbradford.nhs.uk/parents-families/born-in-bradford-s-better-start-cohort.

The Core Outcomes in Anal Cancer (CORMAC: Renehan, Williamson, NIHR) project has pursued NIHR Clinical Research Network adoption; systematic reviews; interviews; a Delphi exercise and plans for a consensus meeting in 2017.

Actionable Healthcare Analytics

HeRC is integrated with multi-partner regional health data initiatives, for example in the Greater Manchester Devolution programme, the centre’s researchers provide leadership in driving the informatics strategy for Health Innovation Manchester, which is underpinned by the HeRC Trusted Research Environment (Ainsworth, Buchan). HeRC led-research into learning health systems methodology attracted an entire edition of a journal, with commentators from around the world contributing articles around our main paper: methods.schattauer.de/en/contents/archivestandard/issue/2286/manuscript/24912/show.html. This methodology underpins the Connected Health Cities programme across North England and the Government policy aligned to it.

HeRC conceived and led the writing of the Health North Connected Health Cities proposal that resulted in a £20m pilot programme to seed the world’s first measurable ‘learning health systems’ in four city regions across North England from 2016-2019 (Buchan, Ainsworth, Taylor). Launching in May 2016 under the aegis of Department of Health, the pilots link Farr Institute methodology with analytic hubs called "Arks", working on data from HSCIC and local NHS and local authority databases to optimise care across two pilot pathways in each region/system.
HeRC led the science components of a successful NHS Testbed application (van Staa; Peek) in Heywood, Middleton and Rochdale working with Verily (formerly Google Life Sciences), MSD and the Greater Manchester Academic Health Science Network to help healthcare professionals better identify and support patients at risk of long term conditions using advanced predictive methods and algorithms: www.england.nhs.uk/2016/01/embracing-innovation.

HeRC’s strategic partnership with the NIHR Greater Manchester Primary Care Patient Safety Translational Research Centre (NIHR £6M; Esmail) has delivered innovations in electronic audit and feedback enabling GPs to identify patients for whom there are missed opportunities for better care, including an electronic medication safety dashboard that is undergoing trials in Salford. A second intervention aims to identify chronically ill patients with actionable shortcomings. Early versions are being tested with GPs (Brown, Wellcome Trust Fellow). PhD student Paolo Fraccaro won the British Computer Society’s 2015 Young Health Informatician Award for his work on this project.

Efficient Trials
Wheat trial. Van Staa established collaboration with Gale (Imperial) with an honorary position at HeRC to deliver this pilot trial, which will enroll premature babies at 20 sites to evaluate the effect of different feeding times around blood transfusion, testing the effects on necrotising enterocolitis. The Farr Institute will provide informatics and analysis expertise.

GetReal project (IMI, development of methods for simple trials). Publications have been submitted on novel study designs (van Staa). Alex Pate received a MRC DTP PhD to extend and generalise his work on this project.

HeRC’s trials research led to the Health Research Agency initiating a public consultation of the expectations of consent in low risk trials (van Staa, Hassan). The report on this public dialogue was recently published (www.sciencewise-erc.org.uk/cms/assets/Uploads/HRA-IRPHR-Discussion-Report-Final-June-24-3.pdf), highlighting the broad support for simple, risk-proportionate consent. The Agency is now finalising revised guidance based upon this, due 2016.

Addressing the hurdles and ever increasing costs of trial regulation, an international consortium of health researchers in 19 countries (including Smeeth and van Staa) wrote an open letter (moretrials.net) to ICH and the EMA expressing concerns over the proposed E6 (R2) “Integrated Addendum” to the ICH E6 Guideline for “Good Clinical Practice”.

The joint MRC Hub for Trials Methodology Research/ Farr Institute Health Informatics Working Group, (Williamson, Appelbe) is running two streams of work; the potential of Electronic Health Records to improve feasibility and recruitment and the use of modern mobile technologies in trials.

New Initiatives
Recognising growing UK strength in veterinary health informatics and potential synergies between animal and human health data science, researchers at the Universities of Liverpool (Radford, Jones, Noble, O’Brien) Bristol, Nottingham Royal Veterinary College, Animal and Plant Health Agency, Animal Health Trust and HeRC are developing ‘Farr@Vet’ to share methodology, training and ‘one informatics’ research (e.g. in zoonosis, antimicrobial resistance and clinical text-mining). A large multi-funder grant (BBSRC/MRC, Radford) application has been submitted and the Inaugural Farr@Vet Symposium was held in Manchester 9-10 March 2016.

By invitation, HeRC is leading theinformatics in a bid to H2020 (UNICORN, Custovic, Buchan, Couch €10m) to provide distributed data analysis over a network of 40 birth cohort studies to better research asthma at EU-scale.

A group institute-wide and beyond have submitted a bid to the CRUK Catalyst Award programme (Renehan, £3.5m). The CREATE consortium will comprise three deaneries; Manchester, Scotland and Yorkshire with the vision to create a sustainable internationally-leading research environment of cancer data scientists going from ‘bench’ (real-world cancer-relevant data) to ‘bedside’ (rapid benefits for patients) in a powerful network of analytics of early cancer diagnosis and population-level treatment optimisation.

Events
George Freeman MP, then Minister for Life Sciences, led a delegation from the Department of Health visiting Vaughan House to officially open The Farr Institute’s Manchester premises on 7 March 2016. The party met researchers, saw demonstrations of science and engineering and explored horizons in health informatics and data science.

HeRC provided the UK keynote for the NIH/NSF Transatlantic Data Science workshop in Bethesda, 1-2 March 2016: sites.google.com/site/usukhealthdata/home

The Farr Institute International Conference 2017 has secured an exciting and ambitious collaboration with Medical informatics Europe for a combined estimated delegation of 1,000 attendees to be held in Manchester April 2017.
Farr London, Led by University College London.

Leveraging the Digital Science Corridor

The Farr Institute, The Crick Institute, The Alan Turing Institute (and the train to the Hinxton Campus) are co-located along a short stretch of the Euston Road.) This corridor is already providing important opportunities for the kinds of inter-disciplinary collaboration which are critical for The Farr Institute’s mission. For example, the first grant secured by The Crick Institute was in partnership with UCL, The Farr Institute and four other partners (EBI, Sanger, King’s, LSHTM, QMUL). This £8m MRC Medical Bioinformatics award (eMedLab) has been transformative not only for providing one of the largest biomedical cloud computing environments in Europe, but crucially because of catalysing new collaborations around health records, genomics and imaging. 2016 has seen eMedLab go live; providing one of the largest biomedical cloud computing environments in Europe, but crucially because of catalysing new collaborations around health records, genomics and imaging. Whereas the UK has an established international reputation for using structured national EHRs for research using primary care records linked with other sources or in using national quality improvement registries, the UK has not had a globally-leading research profile in the exploitation of rich hospital data for research. London seeks to be at the forefront in changing this. Two examples of advancement in ‘unlocking’ real time and imaging data in 2015/16 illustrate the ambition. First, the NIHR and Farr Institute-supported Health Informatics Collaboration. UCLH led on critical care medicine which is widely seen as the most successful of the themes (led from the five major Biomedical Research Centres across Oxford, Cambridge, King’s and Imperial) because it has succeeded in establishing the governance framework and section 251 for data sharing initially across five hospitals, now expanding nationally. Real-time intensive care data collection of about 300 parameters per patient (Brealey, Singer) is considered by many as a culture-changing set of achievements, with few parallels in the world, and forms the basis for ambitious efforts in precision medicine, trial and decision support research. The second example, whether in cardiac or neuro, is that London hospitals perform greater numbers of imaging scans than in most other centres across the world. The question has been whether they can be unlocked for research. Now, in neurology, Nachev has accessed the detailed text reports from 100,000 brain scans carried out at Queen’s Square and developed NLP algorithms for feature extraction which are proving prognostically valid. This work is being extended across London (Dobson).

Major Research Hospitals

Whereas the UK has an established international reputation for using structured national EHRs for research using primary care records linked with other sources or in using national quality improvement registries, the UK has not had a globally-leading research profile in the exploitation of rich hospital data for research. London seeks to be at the forefront in changing this. Two examples of advancement in ‘unlocking’ real time and imaging data in 2015/16 illustrate the ambition. First, the NIHR and Farr Institute-supported Health Informatics Collaboration. UCLH led on critical care medicine which is widely seen as the most successful of the themes (led from the five major Biomedical Research Centres across Oxford, Cambridge, King’s and Imperial) because it has succeeded in establishing the governance framework and section 251 for data sharing initially across five hospitals, now expanding nationally. Real-time intensive care data collection of about 300 parameters per patient (Brealey, Singer) is considered by many as a culture-changing set of achievements, with few parallels in the world, and forms the basis for ambitious efforts in precision medicine, trial and decision support research. The second example, whether in cardiac or neuro, is that London hospitals perform greater numbers of imaging scans than in most other centres across the world. The question has been whether they can be unlocked for research. Now, in neurology, Nachev has accessed the detailed text reports from 100,000 brain scans carried out at Queen’s Square and developed NLP algorithms for feature extraction which are proving prognostically valid. This work is being extended across London (Dobson).

Major Industrial Partners

As well as the usual consultancies and projects we have been devoting more time to the development of strategic partnerships with greater impact. For example, Google has stated that 100,000 people die each year because of the failure of the NHS to mine its records. UCL Investigators Montgomery and Rees have been partnering with Google DeepMind (announced in February 2016) in efforts to improve patient safety using an initial example clinical decision support systems aimed at the management of acute kidney injury using analyses of change in clinical laboratory measurements. This collaboration illustrates scale (20 million blood results were analysed in development of algorithms), pace (two months from research based algorithm to implementable front end for clinicians at the Royal Free) and evaluation for impact (underway).

London as a Global Destination

For London to act as a global gateway for research, education and enterprise in data sciences in health, significant integration activities are required. The Farr Institute has led the way for example in driving a Trusted Research Environment (Ray) which seeks to build on The Farr UCL ISO27001 data safe haven, to provide an environment where, across London, trusts can share data for research and service improvement.
Research
London is developing programmes of research in discovery science, precision medicine, learning health systems, public health and citizen driven health. These programmes are underpinned by: excellent secure data infrastructure; adding value to key linked data resources to make data research-ready; development of multidisciplinary methodological expertise; engagement with patients, the public and health services; strong local, national and international academic partnerships and excellent multidisciplinary training programmes.

Discovery Science and Precision Medicine
Aims to realise the vision of humans as the next experimental model to identify new therapeutic, diagnostic and preventive targets and to understand disease mechanisms. National disease phenotyping efforts through CALIBER build tools to enable phenome-wide approaches to drug repurposing, use of Mendelian randomisation approaches for drug target validation and opportunities for drug repurposing (including research suggesting that glitazone may protect against Parkinson’s disease but showing that angiotensin receptor blockers do not reduce risk of dementia). Translational approaches to drug life cycles using genomic, EHR and trial (e.g. Cochrane Collaboration) evidence. Elucidation of interactions between diseases (e.g. identifying the main initial presentations of cardiovascular disease, showing that COPD increases risk of acute kidney injury; increase in the risk of acute cardiovascular events following Herpes Zoster infection).

Public Health
In 2015-2016, Farr London investigators used EHRs to inform a ‘next generation’ of Global Burden of Disease estimates for England. (Newton et al. Lancet 2015) demonstrating how, for the first time, high resolution, large scale linked EHR resources could be used. Farr London investigators have used EHRs to investigate the health of underserved populations including ethnic minority groups, migrants, homeless people and children in care (Aldridge et al. including Lancet prize for young investigator 2016.)

Learning Health Systems
New initiatives (beyond the Health Informatics Collaborative discussed above) include:

- Collaboration with the University Hospitals Birmingham integrated research and clinical records programme
- EHR4CR; developing EHR standards and interoperable platforms for research in Europe
- Collaboration in European Medical Information Framework (EMIF)
- Learning in Primary Care
- Farr investigators are leading a series of randomised controlled trials (PATHWAY) to improve management of hypertension. These include automated electronic monitoring of home blood pressure as the primary outcome.

- The East London Data Linkage Group have shown the effectiveness of a wide range of multifactorial primary care interventions including managed primary care networks, clinical decision support, education and feedback leading to marked improvements in cardiovascular care, COPD and management of atrial fibrillation).

Citizen Driven Health
Current exemplars include a NIHR-funded trial of Video Observed Therapy (VOT) using smartphones to support adherence. London are also leading a WHO working group to develop target product profiles for VOT which could play a key role in controlling tuberculosis internationally. The CARE CITY NHS testbed for digital health interventions involves partnership with a wide range of tech companies with an overall aim of supporting elderly populations independently in the community.

Decipher My Data!
A research project conducted by Farr London through scientific engagement with schools, promoting the idea that students benefit from performing real science using real data in the classroom. Findings from Decipher my Data! Flu!, based on data from 27 schools, show that school sickness absence data are correlated with influenza surveillance data and may provide a good warning system for spotting flu outbreaks.
The Farr Institute, Scotland has produced 50 publications, achieving over 100 markers of esteem, built a variety of industry partnerships and secured over £10m in research grants over 2015-2016. All of this work contributes to our vision to maintain our vibrant, preeminent, interdisciplinary eHealth Informatics Research Centre in Scotland which allows us to operate as an integral part of the UK-wide Network.

We are ensuring a strategic, collaborative and leveraged approach to the use of public sector data. In 2015 we established the Scottish Informatics Linkage Collaboration (SILC) which is committed to ensuring that citizens’ privacy is maintained whilst simultaneously ensuring that research which is in the public interest is supported. SILC is a dynamic collaboration which allows new partners to join, bring new expertise and to leverage benefits. Partners in SILC include the Scottish Government, National Records Scotland and all academic and NHS partners in the Farr Institute. It is co-chaired by Mr Tim Ellis, the Registrar General, and Sir Harry Burns, the former Chief Medical Officer. It has single oversight of the governance infrastructure, estates and technical/computer infrastructure necessary to support nationwide cross-sectoral data linkage studies.

We launched a series of Frontiers Meetings which are a means to encourage scientific dialogue and collaboration across the UK. We wanted our Frontiers Meetings to be responsive and light touch, supporting new and existing research. So far we have hosted various themes including Imaging, Pharmacoepidemiology, Cancer, Asthma, Kidney Disease and Pregnancy with upcoming meetings in Oral Health, Substance Misuse and Risk Prediction. This activity has resulted in new collaborations between academia, NHS, industry and policy makers, taking forward multi-partner funding applications.

**Trusted Research Environments and Policy Influence**

The Scottish Government published a ‘Safe Haven charter for Scotland’ in December 2015 that builds upon the work of The Farr Institute and Scottish Health Informatics Programme. This charter sets out the agreed principles and standards for the routine operation of Safe Havens in Scotland where data from electronic National Health Service (NHS) patient records can be processed, linked with other data and analysed to support research when it is not practicable to obtain individual patient consent while protecting patient identity and privacy. It also describes, at a high level, how Safe Havens will work together across Scotland on collaborative research projects as part of a federated network - one national and four regional.

**Governance**

The new national analytic platform went live in December 2015, hosted by EPCC and the Advanced Computer Facility at the University of Edinburgh. The new platform has received approval from NSS to hold NHS Scotland data, has been further reviewed under the GCHQ Certified Cyber Security Consultancy (CCSC) scheme and found suitable to hold Governmental data up to ‘Official (sensitive)’ level. This accreditation will allow the platform to undertake research projects involving deidentified HMRC and Department of Work and Pensions data.

A new unified system of Information Governance for NHS Scotland data was introduced. Under this new system the CEOs of Scottish Health Boards have delegated decision making powers to a new Public Benefits and Privacy Panel (PBPP). This panel, launched in February 2016, comprises members of the health service, The Farr Institute research community and lay people. A proportionate risk model is deployed involving two tiers. Tier one meets fortnightly reviewing all applications with tier two meetings held quarterly to consider more complex cases.

**Datasets**

**Scottish National Clinical Imaging Dataset**

We are in the process of obtaining a historical copy and ongoing feed of the Scottish National PACS Clinical System comprising of a range of ‘modalities’ (computed tomography, magnetic resonance imaging, ultrasound, nuclear medicine and plain film radiography) from all 14 Scottish NHS boards. This includes 23 million different examinations from a population of 5 million people and over 850TB of data collected since 2006. So far a prototype software suite has been developed to automate the loading, anonymisation and provision of images within the Safe Haven environment. This is being tested within a regional Safe Haven with the aim to set up the system on the National Safe Haven. The software has been developed as an extension to the Research Data Management Platform (RDMP); a state of the art, scalable, software suit for the management of linked data for research (Jefferson Dundee, Pavis NHS Scotland, Parsons, Edinburgh).

**GP and Laboratory Data**

We aim to facilitate researcher access to de-identified but linkable GP and laboratory data. In both cases it has been decided that it is advantageous to work closely with already existing national level programmes. Accordingly, we continue to work closely with the NHS Scotland SPIRE (GP data), and Infection Intelligence Platform programmes (IIP). The SPIRE programme will undertake a public consultation process in late May and June 2016 (post Scottish election) to ensure avoidance of the issues experienced by care.data. The first laboratory data is expected to be available via IIP in late 2016.

We have prepared and published metadata for our new national primary care prescribing dataset and are currently applying natural language processing (NLP) techniques to enable coding and access to dose instruction data.
Health Board-Level Deeply Phenotypic Data

Research into methods for cleaning and transforming data which meet a range of researcher requirements is been undertaken and the methods developed integrated into the Research Data Management Platform (RDMP). These tools have been utilised to clean and standardise the deeply phenotypic longitudinal records of over 100 datasets.

Impact

SILC is hosting over 300 research studies from academic, industry and NHS investigators.

Precision Medicine Ecosystem. In February 2016 the First Minister announced an additional £4M for a ‘precision medicine ecosystem’ to ensure inter-operability between the NHS TRES, The Farr Institute, Innovation Centres and MRC molecular pathology hubs in Scotland.

Scottish Healthcare Associated Infection Prevention Institute (SHAIP): CSO funding to undertake risk modelling and clinical decision support tools for clinicians focussed on infection management.

NHS Scotland Infection Intelligence Platform: An integrated data platform is now informing and evaluating national antimicrobial stewardship programs in Scotland and is a key strand within the SCOTMARAP 2 – Scottish Government policy document (2014).

Cancer Medicines Outcome Programme: A Scottish Government funded program to examine clinical and cost-effectiveness of cancer medicines in the real life setting.

Innovative Health Care Delivery Programme (IHDP): Led by Dr Aileen Keel, IHDP has been established to fundamentally change the way data can be used to drive improvement in health outcomes by fostering new relationships and connecting the NHS, academia and industry. The team will focus initially on cancer data through the development of a programme methodology which can be extended to other clinical areas (e.g. rare diseases and CHD) on a rolling basis.

The Acute Kidney Injury (AKI): Work led by Wellcome Clinical Fellow Sawhney is supporting the NHS England ‘Think Kidneys’ programme to deliver better care for people with AKI. The 2015 publication ‘Acute kidney injury-how does automated detection perform?’ is listed as ‘recommended reading’ for the Think Kidneys Programme.

Aberdeen Children of the 1950’s cohort: Over 300 cohort participants came together to hear about some of the Aberdeen Children of the 1950s study results and impacts, and the next phase of the study which will look at resilience and successful aging. Participants met with researchers interested in recording the important events in their lives and in discussing the participant’s ideas for future studies. https://www.youtube.com/watch?v=OvA5kB9YwTI


Highlights of the Year

- European Medicines Agency: We have secured preferred partner status to conduct rapid pharmacovigilance studies; one of five in Europe
- Glasgow and Dundee regional safe havens were awarded ISO27001
- 66 new PhDs funded by £4M MRC Doctoral Training Programme in Precision Medicine: Co-led by Robertson and Morris at The Farr Institute in Edinburgh and Pell at The Farr Institute in Glasgow
- The Mason Institute won the bid to host the 13th World Congress of International Bioethics Association in Edinburgh from 14–17 June 2016
- Notable visitors: Ruth McKernon CBE (Chief Exec. Innovate UK), Andy Williams (Chief Exec. HSCIC), Martin Severs (Director of Information & Analytics, HSCIC) Jennifer Dixon (Chief Exec. of the Health Foundation), George Freeman MP (Parliamentary Under Secretary of State for Life Sciences)
March 2015—February 2016


Appendices


Fraccaro, Paolo, Benjamin Brown, et al. 2015. 216 Studies in Health Technology and Informatics Development and Preliminary Validation of a Dynamic, Patient-Tailored Method to Detect Abnormal Laboratory Test Results.


Herrett, Emily et al. 2016. “Text Messaging Reminders for Influenza Vaccine in Primary Care: A Cluster Randomised Controlled Trial (TXT4FLUJAB).” BMJ Open 6 (2). http://bmjopen.bmj.com/content/6/2/e010069.abstract


McCarty, Margaret. 2014. “Care.data: Why Are Scotland and Wales Doing It Differently?” BMJ 348. http://www.bmj.com/content/348/bmj.g1702.abstract


www.farrinstitute.org 36
Appendices


Timmis, Adam. 2015a. “Cardiovascular Mortality in the UK: Good News If You Live in the South.” Heart. http://heart.bmj.com/content/early/2015/05/06/heartjnl-2015-307887.short


Glossary

ADRC
Administrative Data Research Centre

ADRN
Administrative Data Research Network

AHSN
Academic Health Science Network - Collaboration between University, & local Hospital & NHS Trust.

Alan Turing Institute
The UK’s National Institute for Data Science undertaking data science research at the intersection of computer science, mathematics, statistics and systems engineering.

ALSPAC
Avon Longitudinal Study of Parents and Children (also known as 'Children of the '90s study')

APIs
Application Programming Interface

ARUK
Arthritis Research UK

AUKCAR
Asthma UK Centre for Applied Research

Barts
Barts and The London School of Medicine and Dentistry

BHF
British Heart Foundation

Biomedical Research Centres
NIHR Biomedical Research Centres are partnerships between England’s leading NHS organisations and universities, 11 NIHR Biomedical Research Centres conduct translational research to transform scientific break-throughs into life-saving treatments for patients.

Bloom Filter
A space-efficient probabilistic data structure, conceived by Burton Howard Bloom in 1970, that is used to test whether an element is a member of a set.

CALIBER
A research platform of linked electronic health records (EHR) and administrative health data from primary care, secondary care and disease registries. Current data sources include the Clinical Practice Research Datalink (CPRD), Hospital Episode Statistics (HES), the Myocardial Ischaemia National Audit Project (MINAP) and mortality/social deprivation data from the Office of National Statistics (ONS).

CIPHER
Centre for Improvement for Populations Health through E-records Research

CLOSER
Cohort and Longitudinal Studies Enhancement Resources: a programme which aims to maximise the use, value and impact of the member UK longitudinal studies

Connected Health Cities (CHC)
Part of the Government funded ‘Health North’ project led by the Norther Health Science Alliance set up the world’s first partnership using large-scale data to drive public sector reform in health and social care across a 15 million strong population in the North of England.

CPRD
The Clinical Practice Research Datalink

CRUK
Cancer Research UK

CSO
Scottish Government Health Directoriate Chief Scientist Office

DAAG
HSCIC Data Access Advisory Group, about to become the Independent Group Advising on the Release of Data, IGARD
## DataSHIELD
An R library that enables the remote and non-disclosive analysis of sensitive research data.

## DH
Department of Health

## DPUK
MRC Dementias Platform UK

## EAGDA
Expert Advisory Group on Data Access hosted by The Wellcome Trust

## EBI
European Bioinformatics Institute, Hinxton, Cambridge

## eHIRC
e-Health Informatics Research Centre

## EHR
Electronic Health Record

## eLab
A secure data sharing platform developed by Farr Institute HeRC to allow researchers at different institutions to collaborate on data sharing and analysis for research.

## ELIXIR
A European research network that aims to orchestrate the collection, quality control and archiving of large amounts of biological data produced by life science experiments.

## EMA
European Medicines Agency–European Union agency responsible for the protection of public and animal health through the scientific evaluation and supervision of medicines.

## eMedLab
A secure data platform to integrate and share heterogeneous data from personal healthcare records, imaging, pharmacoinformatics and genomics.

## EPSRC
Engineering and Physical Sciences Research Council

## ESRC
Economic and Social Research Council

## GeCIP
Genomics Clinical Interpretation Partnership for Genomics England

## GeL
Genomics England

## Generation Scotland
A unique partnership between the Scottish University Medical Schools, the NHS in Scotland and the people of Scotland it is a bioresource of human biological samples available for medical research.

## GoSH
Great Ormond Street Hospital

## GP
General Practitioner

## GWAS
Genome Wide Association Study

## H@PPI
The HeRC Patient and Public Involvement Forum

## H2020
Horizon 2020. EU funded (80 million Euros) research and innovation programme (2014-2020)

## HeRC
Health eResearch Centre, Led by the University of Manchester

## HMRC
HM Revenue & Customs

## HPC
High Performance Computing

## HSCIC
Health and Social Care Information Centre, now NHS Digital.

## ICO
Information Commissioner’s Office

## ICH
The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) is unique in bringing together the regulatory authorities and pharmaceutical industry to discuss scientific and technical aspects of drug registration.
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICONIC</td>
<td>Improving Cancer Outcomes in the NHS through Integrated Computing</td>
</tr>
<tr>
<td>ICONIC project</td>
<td>Infection Response through Virus Genomics. A Health Innovation Challenge Fund project to utilise next generation sequencing technology, in order to (i) stratify therapy (HIV and Hepatitis C virus used as a case example), (ii) guide hospital infection control responses (Norovirus), and (iii) inform surveillance and epidemiological responses to community outbreaks (Measles and Influenza viruses).</td>
</tr>
<tr>
<td>IMI</td>
<td>Innovative Medicines Initiative. Europe's largest public-private initiative aiming to speed up the development of better and safer medicines for patients.</td>
</tr>
<tr>
<td>Imperial</td>
<td>Imperial College London</td>
</tr>
<tr>
<td>Innovation Centres</td>
<td>Scottish Funding Council initiative to support transformational collaboration between universities and business.</td>
</tr>
<tr>
<td>ISD</td>
<td>National Services Scotland Information Services Division</td>
</tr>
<tr>
<td>ISER</td>
<td>Institute for Social and Economic Research, University of Essex</td>
</tr>
<tr>
<td>Jisc</td>
<td>UK higher education, further education and skills sectors' not-for-profit organisation for digital services and solutions.</td>
</tr>
<tr>
<td>King's</td>
<td>King's College London</td>
</tr>
<tr>
<td>LSHTM</td>
<td>The London School of Hygiene and Tropical Medicine</td>
</tr>
<tr>
<td>MATURA</td>
<td>Maximising Therapeutic Utility for Rheumatoid Arthritis</td>
</tr>
<tr>
<td>MCS</td>
<td>Millennium Cohort Study</td>
</tr>
<tr>
<td>Moorfields</td>
<td>Moorfields Eye Hospital</td>
</tr>
<tr>
<td>MoU</td>
<td>Memorandum of Understanding</td>
</tr>
<tr>
<td>MRC</td>
<td>Medical Research Council</td>
</tr>
<tr>
<td>MRC Medical Bioinformatics Awards</td>
<td>Six strategic awards to combine clinical, health and bioinformatics data. Awarded to: Leeds MRC Medical Bioinformatics Centre; University of Oxford Big Data Institute; MRC/UVRI Medical Informatics Centre at the MRC/UVRI Uganda Research Unit; The MRC Consortium for Medical Microbial Bioinformatics led by Warwick; The Medical Bioinformatics partnership led by Imperial College London; University College London Partners, which includes the Francis Crick Institute.</td>
</tr>
<tr>
<td>MSD</td>
<td>Merck Sharp &amp; Dohme</td>
</tr>
<tr>
<td>NDA</td>
<td>None Disclosure Agreements</td>
</tr>
<tr>
<td>NHS</td>
<td>National Health Service</td>
</tr>
<tr>
<td>NICE</td>
<td>National Institute For Health And Care Excellence</td>
</tr>
<tr>
<td>NICOR</td>
<td>National Institute for Cardiovascular Outcome Research—Collects information from UK Hospitals into secure registries established by the Cardiovascular Specialist Societies</td>
</tr>
<tr>
<td>NIHR</td>
<td>National Institute Health Research</td>
</tr>
<tr>
<td>NIHR HIC</td>
<td>National Institute Health Research Health Informatics Collaborative</td>
</tr>
<tr>
<td>NLP</td>
<td>Natural Language Processing</td>
</tr>
<tr>
<td>NSS</td>
<td>NHS National Services Scotland</td>
</tr>
</tbody>
</table>
### Appendices

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSCHR</td>
<td>Office for Strategic Co-ordination of Health Research</td>
</tr>
<tr>
<td>PHE</td>
<td>Public Health England</td>
</tr>
<tr>
<td>QMUL</td>
<td>Queen Mary University of London</td>
</tr>
<tr>
<td>R</td>
<td>is a programming language and software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing</td>
</tr>
<tr>
<td>RCUK</td>
<td>Research Councils UK</td>
</tr>
<tr>
<td>RECORD</td>
<td>REporting of studies Conducted using Routinely collected Data (Journal of Clinical Epidemiology 66 (2013) 703–705)</td>
</tr>
<tr>
<td>SAIL</td>
<td>The Secure Anonymised Information Linkage Databank</td>
</tr>
<tr>
<td>Sanger</td>
<td>The Wellcome Trust Sanger Institute, Hinxton, Cambridge</td>
</tr>
<tr>
<td>SILC</td>
<td>Scottish Informatics Linkage Collaboration</td>
</tr>
<tr>
<td>SMS-IC</td>
<td>Stratified Medicine Scotland Innovation Centre</td>
</tr>
<tr>
<td>STELAR</td>
<td>Study Team for Early Life Asthma Research (based at The University of Manchester)</td>
</tr>
<tr>
<td>Stoller Biomarker Discovery Centre</td>
<td>University of Manchester. Proteomics facility.</td>
</tr>
<tr>
<td>SUMMIT</td>
<td>European research consortium funded under the IMI programme to identify novel biomarkers for the complications of diabetes</td>
</tr>
<tr>
<td>The Wellcome Trust</td>
<td>The world’s largest medical research charity funding research into human and animal health.</td>
</tr>
<tr>
<td>UCL</td>
<td>University College London</td>
</tr>
<tr>
<td>UCLH</td>
<td>University College London Hospital</td>
</tr>
<tr>
<td>UK Biobank</td>
<td>UK Biobank is a major national health resource, and a registered charity in its own right, with the aim of improving the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses—including cancer, heart diseases, stroke, diabetes, arthritis, osteoporosis, eye disorders, depression and forms of dementia.</td>
</tr>
<tr>
<td>UK HDAN</td>
<td>UK Health Data Analytics Network</td>
</tr>
<tr>
<td>UK MS Register</td>
<td>UK Multiple Sclerosis Register</td>
</tr>
<tr>
<td>UK SeRP</td>
<td>UK Secure electronic Research Platform</td>
</tr>
<tr>
<td>U-PGx</td>
<td>Ubiquitous Pharmacogenomics—European research consortium funded under the IMI programme to investigate if the emerging approach of pre-emptive genotyping of an entire panel of important PGx markers is cost–effective and results in a better outcome for patients.</td>
</tr>
<tr>
<td>WECC</td>
<td>Wales Electronic Cohort for Children</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
</tbody>
</table>