Get Real

Real-Life Data in Drug Development

DATA SCIENCE FOR HEALTH AND CARE EXCELLENCE

Harnessing the UK opportunities for new research and decision-making paradigms

abpi
Bringing medicines to life

NICE National Institute for Health and Care Excellence

MANCHESTER 1824
The University of Manchester

Executive summary

Leading academic researchers working in health data science, clinicians, industry leaders, and representatives from research funders and regulatory bodies met at Manchester Science Partnership's CityLab in February 2016. The discussions focussed on current and future capabilities in data science research and the UK's potential contribution to European projects, such as the Innovative Medicines Initiative (IMI), to use data science to improve healthcare and facilitate the development of medicines. Attendees shared their experiences in the field, reviewed opportunities and challenges for the UK healthcare system, and agreed measures to help overcome current barriers and build on the expertise and data resources in the UK, enabling it to become a leading EU hub for data science and health research using real-world data in the future and attract inward investment.

Several initiatives, such as IMI's GetReal, the European Medicines Agency's Medicines Adaptive Pathways to Patients (MAPPS) project and the Accelerated Access Review in the UK, among many others, are currently underway and are driving the need to consider how best to use real-world data in healthcare decision-making. A wide range of projects involving the use and analysis of real-world data for health and medical research are taking place in the UK, but there have previously been few opportunities for key policy stakeholders and researchers in data science to share their experience and build together on existing expertise.

Key objectives of the meeting were to:

- Explore the current challenges in data science and the factors limiting developments and future progress in the field.
- Share ideas of best strategies to move forward, identifying concrete measures that will support the UK to play a prominent role in delivering the health data science research agenda.

The current challenges for UK's healthcare data science research were identified as:

- **Current initiatives focus on the infrastructure for collecting data** rather than understanding the potential use and value of that data.

- **Lack of dedicated resource to co-ordinate and support UK contribution to external initiatives.** For example IMI and Horizon 2020.
- **No systematic strategy bringing data together to rapidly resolve specific national issues,** with many different data collection systems managed by different organisations with long negotiations for access, stringent data governance requirements and no prioritisation of issues that need addressing.
- **Researchers working in separate 'silos'** with little incentive to collaborate effectively or exchange data, ideas and findings.
- **Shortage of data science skills**. Currently not enough people are being trained to use, process and analyse data, and there is also a lack of further training for people working in the field.
- **Lack of communication and clarity from regulators**, HTA agencies and payers on data requirements in submissions.
- **Lack of public and patient engagement** on their data being used for specific research projects or initiatives.
- **Lack of funding** for research using routinely collected data, particularly for methods, and reticence by journals to publish studies using this type of data. Meeting participants reported a lack of support from research funders for translational research using real-world data and studies bridging clinical practice and research, and difficulties in getting these types of studies published in high-impact journals.

The experts recommended the following measures to advance UK's capability for data science research in healthcare:

Theme 1: Build a collaborative environment

- **Improve collaborative working** by developing networks of people across different sectors with an interest in a specific diseases– academia, clinical medicine, industry and regulators – and enabling them to work together. The right incentives should be put in place, at both political and institutional levels, for people to work together and share research.
- **Establish ways to share data and expertise**, such as with an e-Lab that enables sharing of information and knowledge to overcome the current lack of strategies for bringing data together and many different data collection systems. Technology, governance systems and incentives are required to bring data

together and the group considered it important to optimise the interoperability of technology systems, linking systems together to get the most from them.

- **Encourage patient and public engagement and participation** in sharing data for research. Meeting participants considered it essential to show people the benefits of sharing and re-using routinely collected data in research and in improving care. Initiatives should be set up to empower patients to share their data and engage them in research. This should include reporting back to patients on the findings of studies in which they have been involved so they can understand the value of sharing their data. Stories should be built on using data to improve health and the difference this can achieve, and case studies and examples should be shared.

Theme 2: Develop infrastructure, frameworks and knowledge

- **Further work to establish what data assets are held in the UK**, to include those held by NHS Digital, and promote them globally.
- **Establish funding mechanisms and support** for research using routinely collected data. The group considered it was important to engage funders and help them understand the value of this type of research and recognise that research design and analysis will be different to traditional research studies and clinical trials.
- **Develop training and skills in data science**, with top priorities being mathematical and computational skills, including bioinformatics, statistics, data mining, health informatics, health economics and outcomes research. As users of the data, the public and clinical sectors should also be targeted.
- **Agree best research practice guidelines** for studies using real-world data, including an ethics framework that may include technology to achieve dynamic consent and measures to achieve differential privacy, as appropriate.
- **Involve regulators, HTA agencies and payers in clarifying data requirements**. Meeting participants suggested agencies should better communicate the data they will accept for regulatory approval and technology appraisals. They considered it important that researchers are able to have a dialogue with these decision makers around research programmes and data being used. Current regulations should be updated to reflect new data sources and methodological guidance will need to be developed.

- **Develop quality standards** for databases, to ensure data are of high quality. Data reporting guidelines should define how data should be collected, coded and cleaned, and set out measures to check internal consistency. Gold standards should be established for each dataset.

Theme 3: Leverage current infrastructure and initiatives

- **Derive value from the existing data infrastructure** and promote their utility out of the UK on a fee-for-service analytical basis rather than releasing data. This will include systematic evaluation of NHS datasets such as Hospital Episode Statistics and explore how they might include more clinical information and feedback more actively into guidelines and clinical practice. The group considered it important to ensure that people who collect data benefit from feedback and research using the data, so they can see the value of what they are doing.
- **Scale up initiatives that are working well,** such as the Salford Lung Study.
- **Further develop the national strategy and infrastructure for data science**, with initiatives such as the proposal for a new MRC National Institute of Biomedical and Health Informatics.
- **Think globally** and consider how the UK can contribute to international research programmes.

Meeting participants concluded that the UK has an ideal infrastructure in the NHS to develop research using routinely collected data, and growing experience and expertise in data science. With growing recognition of the importance of research feeding into improving clinical practice and changes in the HTA and regulatory environment for the development of drugs and other medical interventions, it was agreed that measures are needed now to improve collaborative working and to streamline the design and implementation of research using real-world data.

Contents

Introduction

“There are a lot of really good initiatives and projects in the use and analysis of data for health and medical research in the UK. But there is currently no real overview of what is happening in the UK or how different programmes relate to each other,” explained Professor Sarah Garner, NICE’s Associate Director of Science Policy and Research, introducing the meeting.

A vision is emerging through the Accelerated Access Review and activities at the EU level of great opportunities for the UK with its capabilities in data science. Professor Garner challenged participants: How do we build those capabilities? How can we coordinate submissions to organisations such as the Innovative Medicines Initiative (IMI) for funding? The aim of the meeting was to get people working in data science in a wide range of settings and different organisations together to share their experiences and to enable key stakeholders to brainstorm on how to build on the expertise and plan for future developments in the UK. The objectives being to to improve the healthcare of the population and maximise the potential of the UK’s contribution to EU initiatives.

“The key underlying question for today is how the UK can become a global leader in this field?” she explained, with key objectives of the meeting being to:

- Explore the current challenges in data science.
- Share ideas of best strategies to progress, identifying concrete measures that would move this area forward.

Part 1: Data science for research and decision-making

Professor Sarah Garner, Associate Director, Science Policy and Research, NICE

Setting the scene for a shared understanding of data science, Professor Garner explained that it is an interdisciplinary field that brings together processes and systems to extract knowledge or insights about data in various forms. The different types of data emerging in this field include:

- **Big data:** a broad term for data sets so large or complex that traditional data processing applications are inadequate, such as social media data. It is often very unstructured data and may not have been collected for any purpose.
- **Real-world data:** an umbrella term used in the life sciences referring to data that is collected outside standard randomised controlled trials (RCTs), for example pragmatic RCTs and observational studies. It tends to be more structured than big data, and has more in common with epidemiological data.
- **Structured data:** which has 'data models' with data residing in a fixed field within a record or file, for example relational databases and spreadsheets.
- **Unstructured data:** which has no data model or organisation, for example email or the 'notes' pages of documents.

There is a great deal of expertise in the design and analysis of traditional research, but less in the emerging area of the use of computer science in data analysis and bringing different disciplines together, suggested Professor Garner. "At the moment there are lots of disciplines and different fields and they are all talking very different languages. The challenge for us is bringing them together so we have a shared understanding and a shared paradigm," she suggested to meeting participants.

Why look at this issue now? The interest in using these types of data is being driven by new technology being developed and applying that to health. The Institute of Medicine suggested at a roundtable in 2007 that healthcare systems are based on science that provides information that is translated into evidence and this is then translated into care. NICE is interested in this issue as receivers of evidence who then have to translate this into recommendations for care. The challenge with a lot of data is distinguishing signals from noise.

"This is the vision of the learning healthcare system that we want to get to," Professor Garner explained. Information from patients, clinicians and communities inform the healthcare system, which collects and generates evidence and analyses it before it goes back into the healthcare system to inform care. More data and evidence are then collected, which inform care once again. "We have parts of this system in place, with a lot of really good initiatives. Conceptually this is straightforward but the reality is far from easy," she suggested. The process essentially involves collecting raw data, cleaning it up and then carrying out

exploratory data analysis to develop models and algorithms to analyse further data, before communicating results and making decisions.

Several further initiatives are currently underway that are driving the need to consider how to use real-world data in making healthcare decisions:

Research Project	Primary Research Focus
The European Medicines Agency's Medicines Adaptive Pathways to Patients (MAPPS) project	Adaptive pathways is a scientific concept for medicine development and data generation which allows for early and progressive patient access to a medicine.
Accelerated Access Review in the UK	The Accelerated Access Review aims to speed up access to innovative drugs, devices and diagnostics for NHS patients.
The Early Access to Medicines Scheme.	The early access to medicines scheme (EAMS) aims to give patients with life threatening or seriously debilitating conditions access to medicines that do not yet have a marketing authorisation when there is a clear unmet medical need.
IMI - Big Data for Better Outcomes (BD4BO)	The IMI2 Big Data for Better Outcomes (BD4BO) programme aims to catalyse and support the evolution towards value based and more outcomes-focused sustainable and therefore better quality healthcare systems in Europe, exploiting the opportunities offered by the wealth of emerging data from many evolving data sources
IMI - GetReal	GetReal aims to show how robust new methods of RWE collection and synthesis could be adopted earlier in pharmaceutical R&D and the healthcare decision making process.
IMI – Open PHACTS	The IMI Open PHACTS project's online data platform is an online; an open access platform that uses semantic web technology to allow scientists to easily access and process data from multiple sources to rapidly solve real-world drug discovery problems.
IMI - PROTECT	The PROTECT project will enhance the monitoring of the safety of medicinal products. It will also contribute to better evaluate and communicate their benefit-risk profile throughout their lifecycle. To this end, innovative tools and methodological standards will be developed.
IMI - European Medical Information Framework	Europe's largest Big Data project on

(EMIF)	health. Aims to improve access to patient-level data through the EMIF-Platform linking up and facilitating access to diverse medical and research data sources.
IMI - eTRIKS	Enter eTRIKS, which aims to create and run an open, sustainable research informatics and analytics platform for use by IMI (and other) projects with knowledge management needs.
IMI – ADAPT-SMART	ADAPT-SMART aims to create a platform where the conditions and feasibility of MAPPs implementation within the EU regulatory/legal context can be discussed openly and also ensure that MAPPs-related work in IMI is well coordinated.
IMI - ADVANCE	The ADVANCE project will develop and test methods and guidelines in order to pave the way for a framework capable of rapidly delivering reliable data on the benefits and risks of vaccines that are on the market.
IMI2 – HARMONY (BD4BO)	HARMONY is a European network of excellence that captures, integrates, analyses and harmonizes big data from high-quality multidisciplinary sources with the purpose of unlocking valuable knowledge on various hematologic malignancies (HMs).
IMI2 – ROADMAP (BD4BO)	Real world Outcomes across the Alzheimer's Disease spectrum.

In the current scenario, a drug or other innovation is tested in phase 2 trials before being investigated in a larger number of patients in phase 3 trials. If safety and efficacy criteria are met at this point, a marketing authorisation is issued for the drug. After marketing authorisation, patients may initially be studied in observational studies or registries while others will be treated with no active surveillance. The adaptive licensing scenario envisages making greater use of observational data, with an early initial license based on promising data but requiring a company to carry out observational and registry studies once the drug is on the market before being granted a full license based on a very different data profile. "There are a lot of merits in this approach. Patients have earlier access, companies have earlier revenue streams. But how will we put this into operation and structure this?" she asked.

Evidence-based medicine has worked with a hierarchy of evidence that has accepted for the last 20 years that RCTs are superior to other forms of evidence.

This is because randomising patients means all confounders are equally distributed between two groups so the only difference is whether they receive the intervention or not. Other forms of data do not have that clarity. A finding could be a spurious effect, or there could be a bias or confounding factor. One particular concern for medicines is confounding by indication, where the sickest patients tend to be the ones receiving a new drug and so have worse outcomes and more side-effects.

Payers and clinicians need an understanding of the effectiveness of medicines, but data from RCTs gives information on efficacy - how a drug performs under ideal settings. NICE has to take these data and assess how the drug will work in a real work setting. In addition to the traditional RCTs, the types of information that could feed in to this decision-making include electronic health records, healthcare data, registry data, genomics and biospecimens data. How are we going to fit all of this together and make sense of it when making decisions?

Professor Garner concluded that it is essential for regulators to communicate with stakeholders on accepting new types of data for decision-making, how to analyse data and assess data quality. “As a decision-making body it feels like we are facing a tsunami of data. We need some help in thinking through what we need to do with it. We really need to understand where we want to go in order to get the right data in the first place.”

Industry perspectives on UK current capabilities

Dr Shahid Hanif, Head of Health Data & Outcomes, ABPI

Building data science capacity and capability, improving communication within the industry and with other key stakeholders, and optimising the health data environment are key strategic priorities for the pharmaceutical industry, Dr Hanif told the meeting. He suggested that a partnership approach to these developments is essential for making progress.

Why is the UK a good place for health data research?

There are several factors that make the UK ideal for health data research related to the structure and organisation of the health service, existing research capability and policies that support research and development initiatives. As a unified health

system the NHS provides an ideal structure for data collection and analysis, with a wide range of data collected from delivery of routine healthcare by health professionals and from patients. The connectivity of the systems offers the potential to link care records and gain a picture of the whole patient pathway.

Use of primary care electronic medical records for research is well established and a mature research culture has developed around this. A large number of patient records are used for research and the availability of longitudinal records provides the ability to follow patients over time. The UK has world-class strength in academic research, investment in informatics and experience in health data research. In addition, the country has very supportive research and development policies, with research and development tax credits and Patent Box, a special tax regime for intellectual property revenues.

What will ensure the UK remains a world leader in health data research?

An ABPI report, published in 2013, set out an industry perspective on the service levels that would help to ensure that the UK remains a location for research based on routinely collected health and other related data. This focused on providing good quality real-world observational and health information data, providing information about the interfaces between different data providers, and ensuring consistent and transparent rules for licensing data. The report also suggested a need to promote the benefits of research based on routinely collected healthcare data and support for alternative data models and coding systems.

Health data opportunities span the whole spectrum of medicines development and healthcare and include information from genomics through to outcome management. Industry has traditionally worked with structured data but there is growing recognition of the wealth of unstructured data, such as information provided on social media and clinical notes. “Companies each have their own strategy to leverage different sources of data,” noted Dr Hanif, but he suggested a coordinated approach could streamline use of these data sources in the future. Key priorities in this are building data science capacity and capability, improving communication between stakeholders and optimising the health data environment.

The Ministerial Industry Strategy Group (MISG) Health Data Programme

This programme is working towards a vision of delivering a single entry point to access catalogued, linkage-ready, routinely collected health and social care data, within a clear and proportionate governance framework. “The aim is to structure access to data and make the process as simple as possible, in addition to communicating to outside the UK,” Dr Hanif explained. “There are lots of data providers with whom we have to link and ask them if we can share data. How can we harmonise the governance process and streamline access to data?”

Objectives of the Health Data Programme include:

- Robust **delivery governance** for operating and delivering the Programme, with governance to steer the collaboration and the products or services and operating models for the collaboration and for the support of the resulting products and services.
- **Harmonised governance processes** and automation where possible, with a ‘single’ place to start applications for requesting data access, a harmonised application form and supporting evidence, and a ‘single’ decision reached regardless of what and how much data are being requested. There should be approaches for linking data sets and disseminating health data once approved and a ‘charging’ model for accessing health data and associated services and processes and approaches should be automated where possible.
- A harmonised **data governance framework** for data sharing that should be principle-based and proportionate for sharing and disseminating health data.
- A **single access point** for researchers. The Health Data Finder has recently been launched by the NHS National Institute for Health Research, providing a web-based portal to browse the catalogue and to find information about governance and processes, as well to access tools to help with research.
- A single and consistent metadata catalogue providing information about health and social care data presented in a consistent structure.
- A **‘virtual’ research environment** that enables people to share learning and insights about health data sets and their (historical) usage in research, and provide a platform with tools and information to support research.

Health Data Finder

This web-based portal has been developed to help researchers find information about UK healthcare data sets that are available for research and direct them to organisations that can assist in managing access to data. It has been developed by the NHS National Institute for Health Research (NIHR) working in partnership with the Clinical Practice Research Datalink (CPRD), NHS Digital, NIHR Health Informatics Collaborative (NIHR HIC) and Public Health England (PHE). Users can browse the metadata catalogue to find health data sets that are available, with new data sets being added over time. One example of a data set held on Health Data Finder is CPRD GOLD, a primary care data set that contains patient registration information and all care events that GPs in practices registered with the system record as part of their usual medical practice.

The data sets that are available as a priority for phase 1 of the programme include:

- Demographic and vital events data, including mortality and birth data sets (held by the Office for National Statistics [ONS])
- Primary care, including the NHS primary care data set (CPRD)
- Secondary care data, at a national level in the Hospital Episodes Statistics (HSCIC) and at a subnational level with detailed records from secondary care systems for clinical interest areas including critical care (NIHR HIC)
- National data collections, with records from NHS diagnostic services (HSCIC); patient outcome measures (HSCIC); the mental health and learning disability data set (HSCIC); and registries, including cancer registry data (PHE).

Further priority data sets planned for phase 2 include: further primary care prescribing data; secondary care data prescribing data such as IMS MAXIMS hospital prescribing; National Institute for Cardiovascular Outcomes Research (NICOR) data sets; research data sets including UK BioBank, the 100,000 Genomes Project and the Million Women Study; and clinical trial data sets provided by the Medical Research Council (MRC) and NIHR as sponsors of clinical trials, working with industry.

Several case studies are already making use of the strength of real-world health data and expertise in the UK. These include:

- psoriasis and risk of chronic kidney disease – looking at the association between the diseases and the implications for therapy development and healthcare policy
- understanding the true clinical and cost impact of medicine use in everyday healthcare – the Salford Lung Study
- diagnosing cancer and the impact it has on survival times, considering the implications for patients and healthcare policy
- the use of real-world data in economic evaluations.

Improving skills in data science

Bridging the skills gap in data science is also essential to ensure that the UK can remain a location for research based on routinely collected health and other related data. Top priority disciplines to develop relate to mathematical and computational skills, including bioinformatics, statistics, data mining, health informatics, health economics and outcomes research.

The Farr Institute of Health Informatics Research has a major role to play in improving education and skills training in data science. In 2012 the MRC brought together a consortium to establish e-health informatics research centres across the UK. A total of £19 million was awarded to 4 centres based at University College London, and the universities of Manchester, Swansea and Dundee. To further strengthen the UK's capability in analysing and linking health data, the MRC invested an additional £20 million into these centres to create the Farr Institute of Health Informatics Research in May 2013. The aim of the Farr Institute, which comprises 24 UK academic institutions and 2 MRC units, is to carry out cutting-edge research, build research capacity and support infrastructure for enabling safe data sharing.

How to meet the challenges in real world health research

There are several key measures needed to meet the challenges posed in research with real-world data. These include:

- Clear understanding of decision makers' evidence needs. This requires coordination between authorities to avoid duplication.
- Supportive legal and healthcare architecture to facilitate access to data, governance, consent and data privacy.

- Expert knowledge: biostatistics, machine learning, semantics and algorithm development.
- The right infrastructure, with analytical and computing resources; interoperability of computer systems and measures to avoid inconsistencies with coding in data capture.
- Leadership, which is already being implemented with the MISG Health Data Programme.
- Collaboration, exemplified by IMI initiatives for approaches and method: ADAPT-SMART, GETReal, Web-RADR and Big Data for Better Outcomes.

Big Data for Better Outcomes (BD4BO)

This initiative within the Innovative Medicines Initiative 2 (IMI2) programme aims to support the evolution towards outcomes-focused and sustainable healthcare systems and to make the most of medical innovation and opportunities offered by large data sets. Themes include:

- Designing sets of standard outcomes and demonstrate value, with sets of target outcomes, clinical endpoints and alignment of healthcare stakeholders on the value of these outcomes.
- Increasing access to high-quality outcomes data, including mapping of sources, methods and tolls for collection and harmonisation, governance and technical standards.
- Using data to improve the value of healthcare delivery by looking at drivers of outcomes variation, best clinical practices and methodologies to predict outcomes.
- Increasing patient engagement through digital solutions, including patient reported outcomes opportunities, profiling patients' behaviour and tools to increase patient engagement.

There have been three recent calls for projects on Alzheimer's disease, haematologic malignancies and a cardiovascular programme and there are many potential UK collaborators. Further projects are planned in multiple sclerosis and patients with multiple diseases. A 'co-ordination and support' action is also proposed. UK prominence in these initiatives and associated funding would be enhanced by seed-funding to enable leadership co-ordination.

Opportunities and challenges of learning health systems

Dr Niels Peek, Director, Greater Manchester Connected Health City, Health eResearch Centre, Farr Institute of Health Informatics Research, University of Manchester

The current research environment has created separate worlds of healthcare practice and research, argued Dr Peek. “We should translate research questions – things that we do not know about health and would like to know – to the world of research, carry out studies that provide actionable knowledge and translate that back to the world of clinical practice. But in reality this does not always work well and the needs of practitioners are often not properly addressed by research,” he suggested.

There are also problems in trying to translate research findings back into clinical practice. Many RCTs have poor to modest external validity in real-world populations. It can also take a long time for research results to be translated into clinical practice by which time it may no longer be valid. For example, a study carried out several years ago with EuroSCORE, used to predict the risk of 30-day mortality after cardiac surgery, showed the score was completely adrift within 10 years of being published, with the score increasing but the risk going down. “We need a system that can change dynamically over time,” he suggested.

Electronic health records (EHRs) provide an opportunity to use data to achieve this dynamic change over time. They are increasingly collected as part of health care, giving unprecedented opportunities to study population health and the effects of healthcare, with real-time disease surveillance, real-world evidence on treatment effectiveness and dynamic prediction of health outcomes. EHRs can also be used as a platform for experimental studies, with pragmatic e-trials, and for translating knowledge into practice using computerised decision support.

A learning health system has been defined as: ‘an integrated health system which harnesses the power of data and analytics to learn from every patient and feed the knowledge of ‘what works best’ back to clinicians, public health professionals,

patients and other stakeholders to create cycles of continuous improvement.'[1] The aim is to learn from a 'virtuous cycle', in which a problem of interest is identified that researchers decide to study. They assemble and analyse experience data before interpreting the results and tailoring messages to decision makers who then take action. Further data is then collected, analysed and fed back into action.

What are the challenges in using big data to generate new knowledge?

One of the challenges of using 'big data' to generate knowledge and change practice is that the process that generates the data is very different to that in standard research. Routinely collected data are highly transactional in nature, recording interactions between patients and healthcare systems in a way that is driven by the purpose of a patient's visit, which can lead to partially unlabelled data. There is considerable variation in coding practice, variable follow-up times and meaningful events, such as stopping medication, are often not explicitly recorded. "Current research tools are geared to the data currently used, which are clean and complete. The challenge is to develop tools for real-world data," explained Dr Peek. He cautioned that EHR data can be used in a naïve way that fails to take account of potential biases from differences in population samples, clinical information technology, coding practices and data cleaning. It is essential that this type of data is analysed appropriately to take account of these factors.

Computerised decision support systems offer one of the most promising ways of translating information and knowledge into practice. Essentially they use appropriate guidelines that are translated into computable evidence statements to develop a reasoning engine that is used to analyse patient data. Translating text from guidelines into computable evidence statements is laborious, but the University of Michigan is currently developing standards for 'Digital Knowledge Objects'. One approach is *Knowledge as a Service*, in which computerised decision support is provided as a web service.

The Connected Health Cities project

This 3-year regional project in the UK is using large-scale data to drive health and social care reform in 4 city regions in the North of England (Greater Manchester,

---

[1] Friedman CP et al (2010) Achieving a nationwide learning health system. Sci Trans Med 2: 57cm29

North West Coast, Yorkshire and the Humber, and North East and North Cumbria). It started in January 2016 and will focus on optimising local services around 2 or more care pathways and will include public health initiatives such as those tackling obesity, alcohol misuse and the wider determinants of health.

Discussion: what are the current gaps and what should the UK be doing in relation to data science?

Working in small groups, meeting participants identified the current gaps in data science in the UK, before they discussed and agreed their top recommendations for what the UK should and should not be doing to build on existing expertise and experience in health research with real world data.

What are the current gaps in data science?

Meeting participants identified several gaps in data science in the UK, including:

- **Lack of strategies for bringing data together**, with many different data collection systems managed by different organisations. The group recommended that the solution is to develop the technology, governance systems and incentives to bring data together whenever possible.
- **Shortage of data science skills**. There are currently not enough people being trained to use, process and analyse data, and there is also a lack of further training for people working in the field.
- **Lack of clarity from regulators** on what is required from research using real-world data. Participants recommended that regulators should be much clearer on what data they will accept.
- **Lack of public and patient engagement** in sharing their data.

What should the UK be doing in relation to data science?

Meeting participants agreed their top recommendations for what the UK should be doing to build capacity in health research with real world data:

Improve collaborative working by developing networks of people across different sectors with an interest in a specific data area – academia, clinical medicine, industry and regulators - and enabling them to work together. The right incentives should be put in place, at both political and institutional levels, for people to work together and share research.

Meeting participants considered it important to establish ways to share data and expertise, such as with an e-Lab that enables sharing of information and knowledge to overcome the current lack of strategies for bringing data together and many different data collection systems. Technology, governance systems and incentives are required to bring data together and the group considered it important to optimise the interoperability of technology systems, linking systems and data together to get the most from them.

Encourage patient and public engagement and participation in sharing data for research. Meeting participants were concerned that lack of public engagement and support for sharing their data was a major hurdle that must be addressed proactively. They considered it essential to show people the benefits of sharing and re-using routinely collected data in research and in improving care.

Group members agreed that uncertainty remains common in many areas of medicine, so it should be assumed that something could be learned from every patient as part of their routine care. However, there is currently no system in place that encourages everyone to take part in research. Initiatives should be set up to empower patients to share their data and engage them in research. This should include reporting back to patients on the findings of studies in which they have been involved so they can understand the value of sharing their data.

Patients should be empowered to share their data, given information on the benefits of sharing and re-using data and included in research planning. Stories should be built on using data to improve health and the difference this can achieve, and case studies and examples should be shared. Further suggestions included enabling patients to be part of a dataset within the medical specialty providing their care and developing an interface for non-data scientists to query data in real time.

Establish funding mechanisms and support for research using routinely collected data. Meeting participants reported a lack of support from research funders for translational research using real-world data and studies bridging clinical practice and research. The group considered that it was important to engage funders and help them understand the value of this type of research and recognise that research design and analysis will be different to traditional research studies and clinical trials.

Increase training and develop skills in data science. There are currently not enough people being trained to use, process and analyse real-world data and there is also a lack of further training for people working in the field.

Apply best research practice, including ethics procedures and planning projects to answer a defined question. Data should feed into research questions to provide information that is used in translation, feeding back into further research. An ethics framework should be developed for research using routinely collected health data or for research as part of clinical care. This may include using technology to achieve dynamic consent and measures to achieve differential privacy, for example where research is of scientific value and uses data that is not individually identifiable, consent would not be required.

Involve regulators in clarifying data requirements. Meeting participants suggested that regulators and health technology assessors should better communicate the data they need for regulatory approval and technology appraisals. They considered it important that researchers are able to have a dialogue with these decision makers around research programmes and data being used. Current regulatory hurdles were established for a different time and different data and should be updated to reflect new data sources.

Additional suggestions for developing research with routinely collected data:

Develop **quality standards** for databases, to ensure data are of high quality. Data reporting guidelines should define how data should be collected, coded and cleaned, and set out measures to check internal consistency. Gold standards should be established for each dataset.

Derive value from the existing data infrastructure, such as CPRD, and systematically evaluate NHS datasets such as Hospital Episode Statistics and explore how they might be used to include more clinical information and feedback more actively into clinical practice. The group considered it important to ensure that people who collect data benefit from feedback and research using the data so they can see the value of what they are doing.

Scale up initiatives that seem to work, such as the Salford Lung Study.

Further develop the national infrastructure for data science, with initiatives such as the proposal for a new MRC National Institute for Health and Bioinformatics.

Think globally and consider how the UK can contribute to international research programmes.

What should the UK not be doing in relation to data science?

Meeting participants agreed that developments in data science **should not focus solely on the development of drugs**. Data science should be used much more widely in health research. Research should also look at care pathways, ask specific questions and resolve uncertainties in care.

It was considered important to **avoid working in separate silos** but instead work collaboratively. For example, clinical practice should not be separated from research.

Do not wait for perfection in data science. Workshop participants considered that it is important to achieve quick wins now and build on achievements.

Ensure national interests are not isolated from opportunities to collaborate internationally, learning from and contributing to international programmes. UK researchers should contribute to global excellence.

Do not forget the end users, including decision makers, clinicians, and, most importantly, patients. Study results should be fed back to those providing and collecting data.

Do not continue with a negative environment for data collection. Instead, delegates suggested the aim should be to move to a situation where collecting data is the norm rather than the exception. They advised against creating restrictive regulations around data access and consent.

Do not move away from centralising data into single databases.

What should the UK be doing in data science?	What should the UK not be doing?
Improve collaborative working Encourage patient and public engagement in sharing data Establish funding mechanisms and support Increase training and skills in data science Apply best research practice Involve regulators in clarifying data requirements Develop quality standards Derive value from the existing data infrastructure Scale up initiatives that work Further develop the national infrastructure for data science Think globally	Focus solely on development of drugs Work in separate silos Wait for perfection in data science Put national interests above opportunities to collaborate internationally Forget the end user Continue with a negative environment for data collection Move away from centralising data into single databases

Summing up the discussion, Professor Garner said, "What is coming out is the need for national collaboration, with people across different sectors working together, meeting and sharing ideas and expertise." She added, "We need to find the best model for achieving this collaborative effort."

Part 2: Data science: moving beyond the hype

Researchers presented case studies of data science projects currently underway, showcasing what can be done and what could be achieved in the future.

Point-of-care trials: where to go

Professor Tjeerd van Staa, Professor of Health eResearch, Farr Institute for Health Informatics Research, University of Manchester

Point-of-care, or pragmatic trials, are carried out in clinical practice to test interventions and determine whether they work rather than how. Randomised trials have been considered the 'gold standard' methodology for providing research data but they have several problems, including the vast amount of bureaucracy and training required for researchers taking part. Research[2] shows that trials mostly have a relatively small number of participants despite often having huge budgets.

[2] Califf RM et al (2012) Characteristics of clinical trials registered in clinicaltrials.gov, 2007-2010. JAMA 307: 1838-47

They frequently have surrogate endpoints rather than hard outcomes and are relatively short-term with limited or no long-term follow-up. The setting is generally separate from clinical practice and trial governance focuses on audit and monitoring rather than evaluating the accuracy of measuring the outcome.

Professor van Staa suggested that recognition of the limitations of randomised trials and the growing wealth of real-world data provide opportunities for developing pragmatic trials. The Salford Lung Study is a recently completed pre-licensing, pragmatic trial that used EHRs to identify eligible patients and follow them up over time. It was carried out in Salford, where EHRs are integrated across hospitals and general practice. More than 2,500 staff were trained in good clinical practice and local pharmacies dispensed medication to patients taking part. The study illustrates how routinely collected data from various sources can be used in a trial, he argued. “I think this is a very important model for running trials.”

Recognising that use of routinely collected data requires new approaches to study design, Professor van Staa reported 2 potential designs illustrated in recent studies. The Statin WISE study of patients with myalgia thought to be associated with statins recruited participants and randomised them to treatment with a statin or placebo for a period of 2 months before they were re-randomised and treated for a further 2 months. At the end of each treatment period patients received a text message asking about muscle pain, which they answered using a visual scale. Results enable findings to be compared within individual patients. “This is an extremely powerful design for certain questions,” he said.

The trials within cohorts (TwiCs) study design takes a random sample of eligible patients from a large observational cohort and offers them the novel intervention being tested. Refusing treatment would introduce bias if intention to treat or per protocol analyses were carried out. The solution is to use instrumental variable analysis to analyse the data that accounts for confounding with treatment refusal, Professor van Staa explained.

Data and methods transparency: the concept of an e-Lab

The Farr Institute is developing an e-Lab, which is a shared environment using a web-based software application where people can share programmes, understanding of methods and data with colleagues in a secure but transparent

way. “We should move away from reinventing the wheel every time we do something,” he said. An e-Lab brings together people with relevant expertise and authorisation, quality assured integrated data and state-of-the-art algorithms.

Summing up, Professor van Staa suggested there is a need to move to a culture where pragmatic point-of-care trials are carried out in healthcare systems. “We can find out what works best in a particular group of patients rather than waiting for years with a clinical trial.” He suggested that there is a need for re-usable data collection platforms with reproducible and transparent analyses, where data can be shared rather than people working separately in silos. Rather than considering routinely collected data as being too messy for research he considered researchers should use better ways of quantifying uncertainty to take this into account. Alternative designs, such as single-patient (N-of-1) trials or TwiCs should be considered. Finally, engagement of patients and simplified consent procedures are essential to move this forward.

Practical challenges of pragmatic trials: IMI GetReal WP3[3]

Dr Iris Goetz, GetReal work package 3 Lead, Eli Lilly

Illustrating how to solve some of the practical challenges of pragmatic trials, Dr Goetz, explained that GetReal work package 3 (WP3) is focusing on peri-launch relative effectiveness studies. The work package was set up to focus only on pragmatic trials because the group felt this would cover a lot of the issues faced in other study designs and the information learned could be transferred. Pragmatic trials aim to obtain data on how an intervention performs in routine practice through a design that randomises health interventions within a diverse patient population and measures a range of clinically relevant health outcomes reflecting usual care.

The aim of the work package is to identify operational challenges, looking at how they impact on the practical feasibility of carrying out a trial and the acceptability by the range of stakeholders. At the same time, on more of a scientific level, the group is looking at the generalisability and bias of trials once they have applied potential solutions to operational challenges. "We want to offer solutions for operational challenges in pragmatic trials, where possible, and to help trial designers be aware of the consequences of their choices and to maximise the pragmatic nature of study design while maintaining operational feasibility," Dr Goetz told the meeting.

Key activities of GetReal WP3 have included literature reviews and interviews with stakeholders to identify operational challenges in pragmatic trials. The group is working to create a structure that describes and links study design features, highlighting operational challenges and their implications and interrelationships in a usable way. Challenges are being grouped into related 'buckets', such as participants, setting, outcomes and monitoring, and then potential solutions are considered.

Challenges in data collection in pragmatic trials

Focusing on challenges in data collection in pragmatic trials, Dr Goetz explained that data collection and management should follow routine practice at the same time as keeping interference with clinical practice to a minimum. Research sites, physicians and patients need to reflect the prescribers and recipients of the intervention in routine care. Options for data collection include electronic case

[3] https://www.imi-getreal.eu/

report forms (eCRFs) specifically created for a particular study, as generally used in RCTs; extracting routinely collected data from EHRs or disease registries that are already implemented on-site or insurance claims and other healthcare databases, as would be used in pragmatic trials; or a hybrid approach in which data routinely collected through EHRs or other databases are merged with additional data collected specifically for the study.

“There is no one right way of doing a pragmatic trial, which is very different to the design of other types of trial,” Dr Goetz told the meeting. “It all depends on the study question and what you want to achieve with the trial data.”

The challenge of using eCRFs is that they collect a pre-defined set of data at distinct time points. These data sets are specifically collected for the trial by trained staff and are validated closely, which means they do not reflect usual care so may lack generalisability. To solve this challenge, the GetReal team suggests that eCRFs should be kept as simple as possible, only asking for information needed to answer the research question. “Limit it to the variables you need because otherwise you impact too much on routine care and the study will no longer be pragmatic,” she advised. Researchers should also consider whether the information could be obtained from other systems.

When using routine clinical and laboratory data for a pragmatic study, these reflect routine care but may lack detail and completeness. They may be accessible only after a time lag, such as with death certificate data, and data may be collected infrequently or be variable between sites.

The output of the programme is a toolbox that helps study teams trying to carry out a pragmatic trial to navigate through the challenges and to be aware of the impact of their decision throughout the study design and implementation on feasibility and generalisability of the findings. The solution is a good understanding of the process of data entry and management, including where they are coming from and the time points at which they are provided, before starting the study, considering what different challenges in the data mean for the analysis.

“Data management is a balancing act between the requirements you have, which may include those from the regulators’ perspective or regarding ethics, and the risk of interfering with usual care,” Dr Goetz suggested. “Industry, in particular, may try to be ‘super perfect’ and making data collection so complex that you don’t reflect what is actually done in usual care.”

Potential solutions to issues with data collection include on-site staff training, considering the intensity of quality checks and dealing with data errors. “You may be able to deal with this better than you think. There are a lot of statistical methods to deal with these data errors. This may mean increasing the sample size but this may be a better way of keeping up the generalisability of the data,” Dr Goetz suggested.

Challenges in data quality

Similar challenges may occur in data quality, where validity and precision of data have to be balanced against generalisability. Pragmatic trials may show higher levels of errors at data entry but the impact of these errors may be reduced with appropriate statistical methods and/or more control over data entry and quality. The solutions to these data quality issues are to ensure that exact data and data completeness needs are defined *a priori* by the whole study team.

Dr Goetz explained, “The whole team should think ‘what do we actually need, what is the problem if data are not complete?’ You have to define this *a priori* to have all the solutions in place.” A small feasibility study can help to clarify data quality needs, as well as other aspects of a pragmatic study. “This can walk you through these different elements of data collection, and can be done quite quickly. It shows you any gaps you may not have thought through beforehand.”

Looking to the future

Several initiatives are underway on improving the quality, connectivity and use of routinely collected data. These include:

- TRANSFoRm
- IMI EHR4CR project
- FDA’s Sentinel Initiative
- e-clinical Forum.

Summing up, Dr Goetz suggested that pragmatic trials are suitable for obtaining real-world evidence on relative effectiveness earlier in drug development. The design of a pragmatic trial is on a continuum between the classic RCT and an observational study, ranging from more controlled approaches using eCRFs to studies using routinely collected data, such as EHRs. Key challenges lie in obtaining a representative sample of patients, physicians and sites and achieving minimal interference with usual care. This can be achieved by clearly understanding the data collection process, including the level of detail and timing of data collection in relation to the data need, and considering data access and privacy issues before starting a study.

Professor Garner commented that working with the IMI consortium had provided good opportunities for different sectors to work in partnership. “IMI is one way to achieve public-private partnership working. MICA, which the MRC funds, is also supporting industry/academic collaborations,” she noted, adding that nearly 10% of grants awarded by the MRC now involve industry partners in some way.

Making sense of big data

Dr James Weatherall, Head Advanced Analytics Centre, AstraZeneca

Defining advanced analytics and big data “through the lens of a pharmaceutical researcher”, Dr Weatherall explained that his department brings together specialists in scientific computing, biomedical and health informatics and statistical innovation to provide support to drug development decision-making using applied data science. The group analyses clinical and health data to help make the best possible decisions about which drugs to take forward, and for whom. He suggested big data in pharma can be considered in 3 main categories:

- eHealth – routinely collected healthcare data, including EHRs, insurance information and data provided by individual using wearable or smart health apps, which help to understanding medicines ‘in the wild’.
- Genomics – next generation sequencing, Genomics England 100,000 Genomes Project and samples from clinical trials, contributing to understanding diseases.
- Online – data from unstructured sources, including social media, patient forums and feedback, and PatientsLikeMe, all helping to understand patients.

Making sense of eHealth data: case studies in health research and clinical trials

Considering the possibilities that can be realised using eHealth data, Dr Weatherall outlined a collaborative study between his group, the School of Computer Science and the Health Informatics Centre, at the University of Manchester. He also provided an example of looking for clusters of patients and new patterns of comorbidities that would not be recognised in any one centre or by one clinician.

“This is a top down view of a very large collection of electronic health records,” he explained. The group used a new hybrid method, semantic similarity and clustering, put together in a way that gave a novel application. Clustering gave one axis related to diabetic disease, one around patients with cardiovascular disease and another related to respiratory disease (see figure 1). “There are some really interesting hypotheses being generated in the space in between. My question would be: what is this telling us? There are clusters of patients that we may not have resolved otherwise if we had not taken a large data approach.” He suggested that the finding raises the question of whether clinical trials should be based on different types of populations with different patterns of comorbidities.

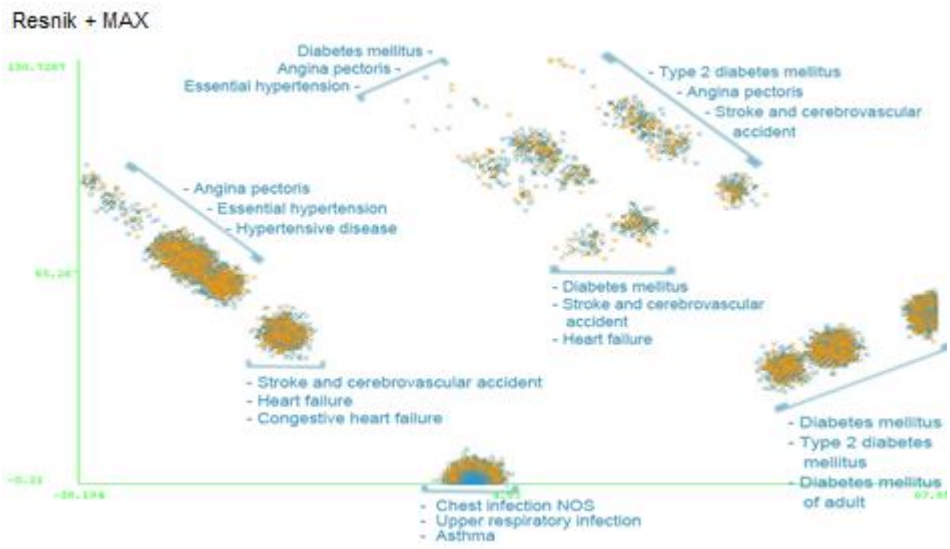
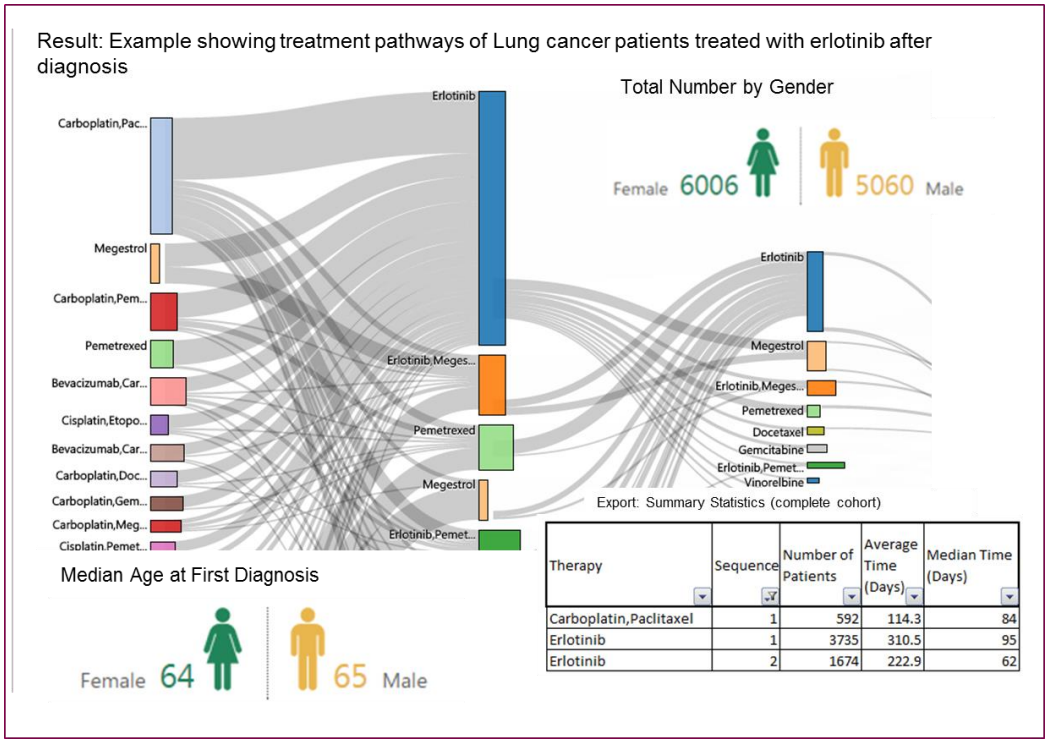


Figure 1: Unsupervised ML: insights into healthcare

Another example is provided by patient flows, such as OncologyFlo. This started with a collection of EHRs used to develop real-world patient pathway mapping. Figure 2 shows a Sankey diagram, representing the flow of patients between

different lines of treatment. The results illustrate what is actually happening in clinical practice. “This approach has helped us simplify an extremely large, dirty and low-quality dataset and find some interesting things about it. The question is: how should this be feeding into the development of new medicines and where are the missed opportunities in these pathways?”



Therapy	Sequence	Number of Patients	Average Time (Days)	Median Time (Days)
Carboplatin,Paclitaxel	1	592	114.3	84
Erlotinib	1	3735	310.5	95
Erlotinib	2	1674	222.9	62

Figure 2: Patient flows – OncologyFlo showing treatment pathways for lung cancer patients treated with erlotinib after diagnosis

Even with messy real-world data a clinically plausible and important result can be found by resolving data at scale. For example, a study looking at liver injury from the use of glitazones in diabetes normalised alanine aminotransferase (ALT) levels at time zero, when patients first received a glitazone. Results showed a statistically significant protective effect of glitazones on the liver by plotting the mean log (ALT) before and after treatment, which showed that glitazone treatment initiation was associated with a fall in log ALT of around 0.15 units irrespective of the other treatments for diabetes. “The approach enabled us to resolve meaningful effects of medication at scale,” Dr Weatherall told the meeting.

The dawn of ‘citizen science’ is also important in eHealth. The evolution of health apps on smartphones and their use as research tools is providing a new way of collecting data for studies. This approach can recruit for trials very quickly. For

example, the Mount Sinai asthma trial enrolled 3,500 people in 72 hours. It is important to consider the balance of speed against representativeness, Dr Weatherall suggested.

There are a huge range of new evidence streams and technologies, including machine learning, EPatients, scalable genomics and digital biomarkers that are going to impact clinical trials in the future. “I don’t know what clinical trials are going to look like in 20 years’ time. But I am sure they won’t be the same as today,” Dr Weatherall predicted. But he felt pressure as an analytics scientist to apply 2 constraints: retain clinical trials as a gold standard and retain a high bar for statistical inference. “Right now, I’m not sure how to solve this equation,” he said, but he described some approaches to help with this.

Device trials

Me & My COPD is a UK digital health support service aiming to enhance the management of COPD by combining mobile devices with integrated medication and symptoms monitoring. Patients with COPD have an inhaler containing a chip that is connected by Bluetooth to their phone, which has a simple algorithm that issues an alert when they should be taking their rescue medication, with the aim of optimising treatment use.

A similar study is being carried out in diabetes in the US where patients are given a web- and smartphone-based diabetes self-management tool. Blood pressure cuffs and weighing scales are connected by Bluetooth to their phone to help them take action to achieve better health outcomes.

Improving agility to respond to data with action

AstraZeneca has developed a ‘watcher’ algorithm for clinical trials providing an interface for clinicians to monitor safety information. Simple algorithms use information streams from a range of different alerts and systems to provide information associated with specific organ systems. “How could this be scaled up to many different information sources so we can pick out actionable information, particularly that to keep patients safe?” he asked.

Summing up, Dr Weatherall suggested the key challenges in making sense of big data are:

- **Contextualisation:** how can we understand new data streams in context?
- **Intelligent aggregation:** we will get benefits of scale but how can we use cognitive computing and fuzzy matching to learn about clinical evidence at scale?
- **Applicability:** rather than being concerned about messy data it is better to consider for what questions is this data fit? For what categories of enquiry does it apply?
- **Agility:** it is essential to ensure that when new evidence streams come on line we surface the information that is really needed to the decision makers who can use it. Where and when is it appropriate to put real-time triggers and alerting in place, as for 'watcher'?

Design issues in trials of digital interventions

Dr Richard Emsley, Senior Lecturer in Biostatistics, University of Manchester

Considering the design issues in trials of digital interventions, Dr Emsley presented 2 case studies of trials of cognitive behavioural therapy (CBT) and psychosocial approaches delivered by digital means to illustrate the challenges and potential solutions.

Case study: using technology to deliver a health intervention - Avatar

Avatar uses technology to deliver a treatment intervention to people with schizophrenia who have treatment-resistant persecutory voices. The client creates a computer image, or avatar, of the person they hear in their hallucinations. The therapist then sits in another room and speaks to the client through the avatar on the computer screen, switching between their own voice and the voice through the avatar. During the course of 6 sessions of therapy the dialogue from the avatar changes from being persecutory to more supportive with the aim of the client feeling they have gained some control over their hallucinations.

A pilot trial with avatar therapy has given very encouraging results. One patient who had suffered auditory hallucinations for about 10 years found they just stopped. The Wellcome Trust invited Dr Emsley's group to apply for a larger trial through their technology transfer scheme.

The typical design for psychotherapy is to take a baseline measurement before randomising patients to a series of sessions of the intervention being tested or to treatment as usual. However, it is becoming increasingly difficult to get funding for these types of trial because the design does not control for non-specific elements of therapy. “We have had to switch to an active control design, where you try to control for non-specific factors such as contact time with a trained professional,” Dr Emsley explained. It could be argued that this is a better test of the ‘active ingredient’ under investigation. If you can show CBT has a significant effect compared with an active control then you are testing the mechanistic effect of CBT, and a mechanistic evaluation can be carried out to show that. However, it has some implications for clinical practice because if there is no significant intention to treat (ITT) effect at the end of the trial there is a question about what this means. Does it mean both treatments were effective or neither is effective?

The Wellcome Trust did not accept the study design proposed, which had treatment as usual as the control, and asked the group to consider how to control for some of the non-specific aspects. They considered this in terms of controlling for 2 possible effects:

- The technology effect of the avatar. The group considered an attention control in which the avatar was created but not used during therapy. However, this had some risks because of the danger of intensifying preoccupation with the voice, so was rejected. Other options were distraction techniques, but there were concerns about harm, and computer-based interactive techniques, but the therapy expectation was poor and so there was an issue with credibility as alternative treatment.
- The therapist encounter effect. One option was a neutral attention intervention but there was doubt about its credibility as a treatment. Supportive counselling was the control invention that was accepted as having been used previously, posing low risk to the participant and probably a low risk to the trial.

The final trial design recommended was either 2 arm or 3 arms, with a treatment as usual group, using befriending/supportive counselling as the attention control. The trial is now going ahead comparing patients randomised to avatar therapy or to supportive counselling. However, Dr Emsley commented, “This is not the design I

wanted because if the trial shows no difference between the 2 therapies we don't necessarily know what that means."

Case study: designing a trial of a digital intervention - Actissist

This ongoing pilot trial is evaluating a mobile phone app that delivers CBT for psychosis (Actissist) for people who have experienced a first episode of psychosis. Users are beeped on their phone or PDA at time points throughout the day and asked a series of questions about their symptoms in that moment, collecting routine data about mental health outcomes that are often missing in care records.

The system asks about 5 domains: suspicious thoughts, voices, getting out and about, feeling criticised and cannabis use. Patients receive feedback as they work through the questions on the app. If they are struggling with a particular aspect they will be linked to information and helpline numbers. Patients receive graphical representations of their data, so they can track their progress, and they have quick access to coping strategies, which can be personalised with each client in advance.

What is the best control for trials with this type of intervention? An MRC-funded trial of an intervention called ClinTouch, which is an interactive monitoring app for people with psychosis provided via a mobile phone, illustrated some of the challenges. In a traditional trial the intervention and control are fixed at the start but the rapid evolution of mobile phones and other technologies over time make long-term follow-up difficult to assess. Trials of digital interventions have to allow for the treatment and the control to evolve over time.[4]

Summing up, Dr Emsley suggested that the development of technology to deliver interventions and to collect measures during routine trials can pose challenges for trial design in choice of control and analysis. "In my opinion, the MRC complex intervention guidelines may not be optimal for these types of interventions because they don't necessarily allow for the nature of the digital intervention and the software to evolve," he concluded.

---

[4] Bucci S et al (2015) Using mobile technology to deliver a cognitive behaviour therapy-informed intervention in early psychosis (Actissist): study protocol for a randomised controlled trial. Trials 16: 404

Final discussion

Professor Garner asked the meeting participants for their top tips to help the UK continue to make the best use of its expertise and resources in data science and play a prominent role in future initiatives, such as IMI. Dr Weatherall suggested a theme throughout the day was that sharing of data would only happen by **engaging the public and patients**. “Somehow I think it has to be a patient movement that is going to be suggesting a lot of the changes we’ve been talking about. Some of the most empowered patients I’ve spoken to are enraged that we don’t make more use of their data.”

Professor Modi noted that it is essential that patients are part of the dialogue on use of routinely recorded data. She reported that the Imperial College London National Neonatal Research Database that contains detailed information on all admissions to NHS neonatal units (approximately 80,000 new patients per year) is used for multiple purposes, including the Royal College of Paediatrics and Child Health National Neonatal Audit Programme. It is populated from predefined extracts from babies’ EHRs as part of patient care, and has strong support from parents. “Parents want us to use these data. They are surprised we are not doing this to a much greater extent. So in the midst of the problems we are grappling with, there are solutions.”

Professor van Staa considered that dynamic consent could be useful, where people can see how their data are being used and can decide if they don’t want their data to be used for particular purposes. “This also allows you to communicate what is being done with the data. We need to be more specific about how we engage the public – detailing specific uses of data.” Delegates discussed the importance of feeding back information to people who have shared their data, noting that companies and academic researchers are increasingly making trial data publicly available.

Collaboration between different people was again emphasised as essential. Professor van Staa suggested that e-Labs provide collaborative platforms where different stakeholders can share data, information and knowledge. “That’s the concept we are working towards rather than people working in silos and

reinventing the wheel." Dr Emsley suggested that, from the perspective of digital interventions, partnership between industry and academia is crucial.

Dwayne Schulthess suggested that **incentives** are also important in encouraging sharing of data. "Right now, the doctor has the data and feels he owns it. How can we split apart these silos and create incentives where people share data?" Professor van Staa suggested one step is to use data to feed information back to the NHS and the healthcare system. "We have tons of data but we don't feed much information back. The first step is for practitioners to realise the value of data collection." He noted that his group is currently working on a 'missed opportunity' study, feeding information back to clinicians where guidelines are not followed. "The feedback is very positive because clinicians put a lot of work into records but don't have time to go back through them."

Dr Goetz added that **practicability** is essential. "Clinicians have no time to do anything other than their clinical work. It's not that they don't want to take part in research but they just do not have time. So we have to create something that is easy to implement," she argued.

Summing up the meeting, Professor Garner said, "The take home message is that this is difficult but not insoluble. If we are going to make progress quickly it will be faster together. We have to break the problem down, assign responsibilities and think collectively about solutions."

Acknowledgements

The meeting organisers would like to thank and acknowledge the meeting participants for their contributions.

Meeting participants

Adeeb Ahsan
Scientific Project Manager, Science Policy and Research Programme, NICE

Yvonne Alexander
Professor of Translational Science, Manchester Metropolitan University

Kate Aylett
Programme Manager, Medical Research Council

Sumeet Bakshi
Senior Director, Consulting and Analytics, LASER

Mike Burrows
Managing Director, GM-AHSN

Mike Chambers
Independent Pharma Consultant

Chris Chinn
Head of Real World Investigations, Sanofi

Vasa Circin
Lecturer in Health Informatics, Kings College London

Nick Crabb
Programme Director, Scientific Affairs, NICE

David Crosby
Programme Manager for Methodology and Experimental Medicine, Medical Research Council

Vincent Doyle
Associate Director, Digital Services, NICE

Mark Edwards
R&D Director, EMIG

Richard Emsley
Senior Lecturer in Biostatistics, University of Manchester

Chris Gale
NIHR Clinical Trials Fellow, Imperial College London

Leslie Galloway
EMIG Chairman, EMIG

Sarah Garner
Associate Director, Science, Policy and Research Programme, NICE

Wendy Gidman
Technical Analyst, ODU Appraisals, NICE

Iris Goetz
Research Advisor, Center of Excellence Epidemiology, Eli Lilly

Shahid Hanif
Head of Health Data & Outcomes, ABPI

Pall Jonsson
Senior Scientific Adviser, Science, Policy and Research Programme, NICE

Farid Khan
Chairman, Elucid

Andy Lawton
Global Head of Data Management, Biometrics & Data Marks, MD, SVP Projects, Clinical Platforms & Science, GSK

Johnny Lundgren
Chairman, GM-AHSN

Anna-Katharina Meinecke
Researcher, Bayer

Fay McCracken
Technical Adviser, NICE

Andrew Mitchell
Senior Information Architect, NICE

Neena Modi
President, Royal College of Paediatrics and Child Health, and Professor of Neonatal Medicine, Imperial College London

Nuala Moran
Real World Data Researcher

Vishal Patel
Senior Manager, LASER ANALYTICA

Niels Peek
Director, Greater Manchester Connected Health City

Chris Pomfrett
Technical Adviser, NICE

Duane Schulthess
Managing Director, Vital Transformations

Ian Shemilt
Senior Lecturer in Economic and Evidence Synthesis Methodology, IOE

Rob Thwaites
Senior Director, Takeda

Alexia Tonnel
Evidence Resources Director, NICE

David Tyas
Associate Director Health Economics & Outcomes Research, BMS

Tjeerd van Staa
Professor of Health eResearch, Health eResearch Centre, Farr Institute for Health Informatics Research, University of Manchester

Tom Walley
Director, NIHR HTA and EME Programmes; Professor of Clinical Pharmacology University of Liverpool

Radek Waziak
Evidera

James Weatherall
Head Advanced Analytics Centre, AstraZeneca

Andrew White
Head of Medicines Optimisation, NWCSU

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no [115546], resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies in kind contribution."