

Digital Health Data: what types of value can be generated, and how do we do it?

Date and Location: 4th December 2019 at The Royal Society

Chair: The Rt. Hon, the Lord Willetts FRS
Chair, The Foundation for Science and Technology

Speakers: Dr Saira Ghafur
Lead for Digital Health, Institute of Global Health Innovation, Imperial College London
Caroline Cake
Chief Operating Officer and Deputy Director, Health Data Research UK
Dr Natalie Banner
Understanding Patient Data Lead, Wellcome Trust
Dr Nicole Mather
Life Sciences Lead, IBM Global Business Service, UK & Ireland

Report Author: Dr Caroline Pritchard
Impact Manager, National Measurement Laboratory, LGC, and Foundation Future Leader

Audio/Video Files: www.foundation.org.uk

Hash tag: #fsthealthdata . Twitter Handle: @FoundSciTech

DR SAIRA GHAFUR, Lead for Digital Health at the Institute for Global Health Innovation, began her talk by highlighting that the English NHS has the most comprehensive longitudinal health data set in the world. This data set has significant potential financial value to the UK; indications from UK Treasury estimate that public sector knowledge assets are valued at £150 bn with a potential return of at least £5 bn per annum. Although patient benefit must remain a key priority for how data is shared and used, there is a real opportunity to gain financial return for the taxpayer. In fact, missing out on public return could result in a loss of public trust in data sharing if, for example, a new digital technology were developed and the NHS was not recognised or compensated for its input.

In order to harness the full potential benefits of this data for improved patient care, a significant investment in resource and capability, as well as infrastructure and capital, is required. With health care providers having to use as many as 1000 different systems, there is more work to be done to improve

the data that exists, updating and improving current data collection systems such they can be used and linked most effectively. In addition, the real costs of creating these NHS data sets, from front line data collection to storage and linking at a national level, must be considered and accounted for in order to better understand their value. Saira highlighted that the human capital cost, e.g. clinicians' time entering data into electronic health records, represents a large fraction of this cost currently unaccounted for.

Progress towards creating more useable data sets is being achieved in part through investment in internal capability, for example in the NHS Digital Academy, but also through reliance on industry partnerships. Saira outlined the different value sharing mechanisms that are open to the NHS, from free/discounted access at one end to an equity share in the company commercialising the data at the other. For example, the partnership between DeepMind and Moorfields Eye Hospital includes free access to retinal images

over the next five years. Although this raises the question of equality, as other Trusts have to pay for the same services. Saira highlighted that when deciding where financial gains from data sharing go, whether to individual NHS Trusts where the data originated, or the NHS, or a separate body (e.g. sovereign health fund as proposed by Parliament), it is vital to ensure this does not create further inequalities in geographies. Whichever partnership mechanism is chosen will be case specific; dependent on the extent of NHS contribution, the work required by a commercial partner to effectively use the data and the type of product being produced. Most likely there will be combination of options that balance risk and reward while protecting a certain return for the NHS. A centre of expertise to support individual NHS Trusts in assessing these options to achieve a fair return for the public is also necessary.

The role of public acceptance and trust in this area is crucial. Events such as the WannaCry attack (2017), Facebook/Cambridge Analytica scandal (2018) and recent Google-Ascension Health partnership data transfer has led to reduced public confidence in data sharing with technology companies (YouGov survey). There must be a clear governance framework within the NHS that addresses the questions of privacy, ethics and security. This framework must consider principles that already exist covering the ethical issues of data sharing, e.g. Human Rights Act 1988, GDPR, OECD Recommendation for the Council of AI. A policy framework has been developing rapidly within the UK over the last year through initiatives such as the Code of Conduct for Health and Care Technology, the proposed launch of Centre of Expertise by Office of Life Sciences, and the launch of NHS X in 2019. As this framework develops, the NHS must prioritise patient benefit while respecting the privacy, ethics and security of NHS patient data to ensure public trust in data sharing.

CAROLINE CAKE, Chief Operating Officer and Deputy Director at Health Data Research UK (HDR_UK), described the role that HDR_UK, launched in 2018, is playing in providing safe and secure access to health data for the benefit of patients. HDR_UK is the national institute for health data science. Working across the NHS, industry, academia and patients, one of HDR_UK's focuses is on creating a robust health data research infrastructure for the UK, as identified in the Industrial Strategy. This initiative will allow greater efficiency of research, more meaningful and effective novel

medicines or treatments, faster commercialisation for business, and increased employment and growth, ultimately leading to improved health care for us all.

To help guide how this new infrastructure develops, HDRUK engaged with thousands of individuals and hundreds of organisations across the UK, to identify the concerns, obstacles and challenges that people, business and researchers currently see in accessing and using health data. Common themes emerged from these discussions. Although there are good data resources in the UK, it is unclear for many researchers and innovators what actually exists and whether (or how) it can be accessed, resulting in innovation and research happening outside the UK. The public do support research using health data but only if they understand how the data is being used and what the benefit to the NHS will be. Caroline also highlighted that it was crucial that the public and patients are involved in the process of data access so they clearly understand the benefits as well as how any risks might be managed, and a Public Advisory Board will help guide decisions within HDR_UK.

Alongside investments in science and building the UK's capability and skills, encouraging data science as a career path, the health data infrastructure will further health data science in the UK in three broad areas: standards, curation and improvement of data, and establishing a single common access point.

The Health Data Research Alliance, established earlier this year (2019), addresses data standards, establishing best practice for the ethical use of UK health data. The Alliance, which consists of a growing number of national bodies, charities, NHS Trusts and researchers, will help to agree standards and quality requirements for data custodians, in order to build trust in the data and its uses.

Health Data Research Hubs address data curation and bring together industry, academia and clinicians to focus on disease specific data sets to improve the secure and responsible access to this data for health research and innovation. Caroline highlighted that the process of creating these consortia has provided invaluable understanding as to how these different organisations work together to reach a common goal. There are currently seven Hubs covering areas such as cancer, respiratory health and clinical trial data, amongst others, which span the whole of the UK.

Finally, the HDR Innovation Gateway will provide a common search point to UK health research data for accredited researchers or innovators. The first phase of this project will provide a Minimum Viable Product to

be launched in January 2020, with a more sophisticated model currently being commissioned for Phase 2.

Ultimately, this infrastructure will help provide what people want: high quality health data from across the UK that provides longitudinal information, is connected from across different elements of an individual's life, is easy and quick to search for and request access to, and provides real-world information. Caroline highlighted that it is imperative that this is underpinned by increased confidence and trust by the public in exactly how data is used to achieve real benefit for patients.

DR NATALIE BANNER, Understanding Patient Data Lead at the Wellcome Trust, discussed the value of data from a societal and individual perspective. The trustworthiness of data is a crucial issue for the public. Any systems or governance structures that are developed must be worthy of the public's trust such that they accept the level of potential risks, however minor, for example lack of privacy or misuse.

The primary question that should be asked whenever we use health data is: what is the purpose of the data? In order to understand whether there should be a financial value associated with the data, it must first be clear what is happening to the data. Data is not intrinsically a property and cannot be owned in the same way an object can, which leads to fears around data use. It is important to understand what the patient and public concerns are, and these go beyond privacy and consent and includes expression of choice over what subsequently happens to their data.

The Office for Life Science and NHS England created public citizen juries to allow the public an unbiased opportunity to explore the questions of what a fair partnership might look like, and what a fair return for the NHS should be. The results have not yet been published but Natalie outlined some of the initial results. The primary concern is that there must be public benefit of data sharing. Natalie highlighted that there is significant scepticism around whether the benefits that are currently being articulated, both financial and those for the patient, are realistic. It is not useful to the future acceptance of health data use to overpromise at this point. Despite the anonymisation and compilation of large data sets, people still feel a personal stake in data and want reciprocity; any benefits must come back in some way to the NHS or the patient community. The public also expressed concerns over further health inequalities within the NHS; will those Trusts that have advanced data infrastructures benefit further from

commercial exploitation. Any commercial exploitation, large financial profits and potential trade deals are all seen negatively by people, although there was a general consensus that the only thing worse than selling off patient data would be to give it away for free.

There are two challenges that must be addressed as the UK moves forward in this area. First is the illusion of objectivity of data. Data itself is not value-free. It may not be truly representative of a population, with particular groups being excluded and resulting in inaccuracies of results that lead to safety concerns. Algorithms that manipulate the data may have inherent unconscious bias, such as the recent Apple Card scandal that showed gender discrimination. The value inherent to data is currently only considered when something goes wrong but should be addressed from the outset. Secondly, it is important to consider who has the ultimately responsibility for deciding what is defined by 'public benefit'. What may be considered of benefit to some groups, particularly those that have more power, may be to the detriment of others. There needs to be a diversity of voices and common questions to help comprehensively examine these issues.

Natalie highlighted that it is important to understand that health data is a means to an end and not the end in itself. There is a potential danger of developing purely technical solutions to the current data challenges. By framing the question differently, asking only what patient-focused challenges could be investigated more effectively if we had better data, can the value of the data truly be maximised. Identifying where and how we can use data most effectively will ensure that the UK is both ambitious but realistic in its use of health data, and develops fundamentally trustworthy solutions for the benefit of patients.

DR NICOLE MATHERS, Life Science Lead at IBM Global Business Service, provided a different perspective on health data. Digital Health is one of the new industries that the UK is taking advantage of under the Life Sciences Industrial Strategy to increase investment within the UK and improve benefit for patients.

The NHS has the components to form rich longitudinal patient data records. Before the existence of Health Data Research, the data existed across many different institutions, in many different types and under different governance structures with limited connectivity. This includes primary and secondary care data, deep records within NICOR, genetic data within Genomics England. Health data has the potential to inform cost effective pathways, explain why products might work and improve understanding of

disease progression, among others.

Nicole provided an example use case of health data in the clinical trials area. An HDR partnership between IBM, Oxford University, Microsoft and NHS Digital aims to improve the assessment of clinical trial feasibility. This system includes patient identification and management systems to allow more flexibility during the trial. Geography no longer becomes a restriction, as patient records from across the UK can be easily accessed through NHS Digital to find suitable candidates that would previously have been missed. With a rich data set, patient controls may no longer be necessary as synthetic controls can be used, resulting in all patients on the trial being treated. With trial outcomes incorporated back into the patient record, this further improves the existing data set. Health data can improve productivity of clinical trials, improve opportunities for patients and researchers.

Data assets are distributed across the UK, between the NHS, technology capability and the research community. Individual companies cannot create value and deliver patient benefit alone, and there is often a significant investment required in making the data that does exist useable. Nicole highlighted that there are a variety of models that can be used for the NHS to capture value from the data: from data sharing agreements, licensing or commercial arrangements, all different options for creating value within the NHS are in use. The Office of Life Sciences have performed a study to investigate these different approaches within the NHS (Making NHS data work for everyone, Reform (2018)). There is a continuum of options that range from free access to commercial IP for NHS product development and Nicole commented that it is helpful that there is no drive to standardise this approach.

However, as outlined in the DHSC Code of conduct for data-driven health and care technology (2018), whichever option is chosen, fair share remains a guiding principle. Nicole highlighted that this will remain central to any organisation in delivering benefit from NHS data; a fair share must go back to the NHS and patients.

In order to deliver value, the data must be accessible, and appropriate capabilities are required to access it. This requires a complex system of data storage, curators and analysis. A significant amount of time and investment must be focused on the ground work, improving the collection and curation of data such that it is aligned and can be effectively integrated to make it useable. As with data assets, the expertise in these areas is

disparate and at each stage consortia are crucial to enable progress.

There is real potential to develop patient benefit from health data but collaborations and partnerships will be crucial to achieving this. In each case, the outcomes of using the data must be clear, and any commercial benefits to the NHS and other participants need to be well-defined and appropriate. Working together in a way that maintains trust in and security of data at each stage, health data can make a real difference.

THE DEBATE

The debate that followed touched on a range of issues. The UK's position relative to other countries was discussed and the challenges that might come from external competition. It was highlighted that the UK has an internationally strong reputation for a good regulatory environment that respects the rights and interests of the various players, and that this is a unique selling point for the UK. There is also value in competition as the breadth of different perspectives on health data systems from across the globe will allow the field to progress collectively for the benefit of all.

Potential risks of using health data were commented on and concerns around whether you could track individuals through linked data that could be used for, for example, insurers or employers to make decisions, were raised. Using the wrong data for the wrong purposes is a real potential risk. No one within the health data eco-system wants patient health data to be used for insurance or marketing purposes. Any risks must be minimised within the governance frameworks that are developed.

These frameworks should be developed in a meaningful way, as it is imperative that the public understand what is happening with their data. Property and the concept of ownership is a poor analogy for data, as data can be licenced or copied and there is limited choice and control once this happens. People do not read terms and conditions and data contains information not only about us as individuals but also others, e.g. genetic data. Privacy should not become a luxury commodity. The current opt-out option has limited scope as it only covers confidential data, such that the opt-out doesn't apply in most cases. It is important to be open about this with patients, to educate them as to how exactly how their data is being used to prevent any shocks or surprises further down the line.

Charities will play a crucial role in the use of health data as the public have a high degree of trust in their work. They can be a gateway to help increase trust in the use of health data in the wider NHS. Recent GDPR changes and negative stories about data use has made people more aware of their data. It is crucial that the health data community protect the relationships of

trust the public have with their GPs around their data use and enhance confidentiality.

The question of who decides on what the benefit to risk balance is was raised, as different people are comfortable at different points on this scale. In each case it is important to understand where on the scale an opportunity exists and balance benefit and return through the development of new companies to the potential for the implementation of poor technology on the front line. There need to be feedback loops in place to identify where errors in judgement and data access have occurred in order that they are not repeated.

However, currently patients can often not even access their own data. There are fundamental infrastructure questions that need to be addressed before any further technology, AI or research applications are considered.

The potential for SMEs to create value and jobs through NHS data was identified but the confusion as to routes to access the appropriate data were raised.

Examples such as DigiTrials, where accessible data for clinical trials is being developed, were given. It was highlighted that it takes time to build confidence around user data and develop well-curated data sets.

The processes that are used on these initial data sets can then be streamlined and integrated. Working with start-ups will provide a route to identify and overcome the challenges of developing accessible data sets.

The cost and value of collecting health data was discussed. Different ways of collecting data have evolved over time, for example electronic health records are now used rather than paper record. It is important to allow time for the adoption of different systems. This will help ease the burden of front line data collection to reduce the initial effort so less curation is required. Health data is moving towards patient-held records to share patient information. Patients themselves can play a critical role in improving front line data quality; highlighting inaccuracies in their own records. This role should be emphasised as access improves both data accuracy and engagement. There is some initial indication that this could lead to enhanced patient safety.

The GP-patient relationship could be further enhanced through the use of data at the front line. This supports tools that are already embedded, though further work with clinicians around developing skills around engaging with algorithms could be done. It allows the patient to see that clinician decisions are evidenced based rather than anecdotally drive. Stronger relationships could provide a powerful opportunity to improve front line decision making and engage patients more closely.

Inequality drivers could have social causes not picked up by data. For example in Wales, there are initiatives to bring data from different sources, education, social

and environment, health. This could be used to answer some interesting public health questions.

The narrative that is used around health data will be critical as without patient permissions and engagement, none of the potential benefits will be possible. The opportunity costs and benefits must be clearly communicated to the public. The potential for using the counterfactual argument was also discussed in some detail, should the public understand what happens if we don't access this potential resource, and whether communicating the value in this would encourage or discourage public engagement. Examples that demonstrated the human cost of not using the data could be considered to convince the public.