How better use of data can help address key challenges facing the NHS

Josh Keith, Fiona Grimm, Adam Steventon

The Health Foundation

Contents

Key points

- This long read explores how the NHS in England can better use routine health data to help address current challenges, including winter pressures, the ongoing coronavirus response and the growing elective care backlog.

- We examine the longstanding barriers to widespread use of data and data science, consider what actions might help to overcome these, and explore whether the data strategy for health and social care will deliver the change needed.

- We know that better use of routine NHS data can directly support care, drive service improvement and enable research and innovation. To realise these benefits, the NHS must address enduring challenges with how the NHS approaches data science and data-driven innovation.

- The NHS needs both a sustained programme of action to tackle the underlying barriers, and much closer alignment between efforts to improve the use of data and other efforts aimed at transforming health and care. We propose five areas of focus:
    - improving underlying infrastructure for data and technology to provide high-quality, timely data for service improvement, research and innovation
    - developing the analytical workforce and better harnessing their skills
    - focusing on data-driven innovation as a service, with routine development and deployment of open-source innovations developed in collaboration with end users
    - building better implementation infrastructure accompanied by effective regulation, monitoring and evaluation, to ensure safety and equity and to build confidence among health care professionals and the public
    - fostering a responsible approach to innovation to ensure everyone's health care benefits.

- Changes driven by the NHSX data strategy and the Wade-Gery review are important steps in the right direction, but there is room for improvement. The chances of the final data strategy achieving the change needed could be increased with a clear and credible delivery plan, and by addressing parts of the strategy where current commitments do not go far enough. Such areas include how data are used to understand and address health inequalities, and the need for a more explicit commitment to engaging the public to help ensure that data use is worthy of people's trust.

- Beyond national-level alignment of efforts on data, digital technology and transformation under NHS England and NHS Improvement's 'transformation directorate', local NHS organisations – who are critical to delivering change – need further clarity on why better use of data is important, what they can do, and where funding and other support will come from. Success also depends on action from national health care leaders, managers, clinicians and analysts at all levels of the NHS, and their collaborators in academia and industry.

Introduction

As the NHS continues to deal with the pandemic, there is also huge pressure to find better, and more efficient, ways to deliver care and reduce backlogs, and to target health services and prevention activities at the groups who are most in need. Tackling these challenges will require the NHS to make use of all the tools at its disposal.

Data and data-driven innovation could be a part of the solution. A number of strategies (the data strategy for health and social care, the national AI strategy for health and adult social care) and legislative changes (the Health and Care Bill) seek to deliver the platform for such an approach. The Wade-Gery review has also resulted in significant reorganisation to support the transformation of health and care.

In this long read, we first consider the contribution that routine health data can make to addressing some of the challenges faced. We then move on to examine the longstanding barriers to widespread use of data and data science, and what approaches might help to overcome these. Finally, we explore whether the data strategy will provide much-needed coherence to this area.

1. How can health data help the NHS?

The potential to derive benefit from routinely collected data[1] sets the NHS apart from many other health systems globally. The NHS is a universal, publicly-funded health care system covering a population of 56 million people in England. The NHS could learn from every patient interaction to continually improve services, better understand health and care needs, develop new treatments, support advances in data-driven technology and AI, and enable more efficient and patient-centred care. Furthermore, the single-payer nature of the NHS, along with past investments such as the introduction of the NHS Number, allows for rich longitudinal data. Such data are important for understanding the causes of ill health and assessing their risk. Table 1 sets out how good use of health data is also critical to responding to the challenges currently facing the NHS.

---

[1] The data generated, collected and stored by the NHS as part of the delivery of services. This includes data from every contact between patients and the NHS, and wider administrative and clinical data such as diagnostic test results.

Table 1: Potential contribution of data and data-driven innovation to NHS challenges

Challenge	Examples of the impact of data and data-driven innovation
Addressing the growing backlog and waiting times	Data and data-driven tools can support improved delivery and increased efficiency and capacity of outpatient services, and reduce time to diagnosis and treatment.
Addressing widening health inequalities and improving prevention through personalised and anticipatory care	Novel analysis and better linking of data can help NHS commissioners and providers measure inequalities (unequal access or outcomes, for example), understand their causes and allocate resources more equitably. Data can also enable anticipatory and personalised care by supporting the identification of at-risk populations.
Accelerating the adoption of effective, evidence-based service innovations	Rapid evaluation using routine data can help assess the impact of service innovations such as digital first primary care. New approaches to rapid research and evaluation, such as those being pioneered by the Improvement Analytics Unit and Thiscovery, can enable this. Patient-facing data-driven innovations also have the potential to help people better manage their own health.
Reversing stalling improvements in health outcomes	Good use of data is central to delivering improvements in the quality of services and patient outcomes. It can be used to identify areas of poor care, guide choices about priorities for care, help develop interventions to improve efficiency and quality of care, and evaluate efforts to integrate care.
Ensuring ambitions are supported by funding, resources and plans, including for the workforce.	Good quality data are critical to developing the evidence needed to inform long-term funding plans, for example through modelling the future demand for care, the workforce and other resources needed, and to understand the impact of reforms.

The case studies below set out what this contribution could look like in practice.

Case study 1: AI-driven clinical decision making to manage cystic fibrosis

Chronic conditions such as cystic fibrosis (CF) make up an increasing proportion of the work done by the NHS. Current management of CF requires routine, specialist-led check-ups approximately every 1–6 months for adults. This approach is costly to the individual and the NHS, and exposes patients to additional risks of cross-infection in hospital.

The Health Foundation, in partnership with Health Data Research UK, supported Project Breathe, led by Royal Papworth NHS Foundation Trust. Project Breathe sought to transform the delivery of care through Bluetooth home monitoring devices, and the development of an AI-driven approach to predicting future flare-ups for people with CF.

The project showed that the digital solution could reduce the volume of routine CF outpatient appointments by 50%, and support patient empowerment and self-management. There is also early evidence that the AI-driven approach can predict the development of pulmonary exacerbations up to 10 days before they become clinically apparent. The work has been scaled to three further CF specialist centres in the UK, and the predictive algorithm is now being taken forward for clinical trial to assess safety and efficacy.

Case study 2: Improving safety and patient flow in intensive care

The intensive care unit (ICU) is a complex environment. Patients can be admitted to the ICU from, or discharged to, many different parts of a hospital. Demand for ICU beds is growing, and patients have more challenging, unpredictable care needs, making it difficult to optimise resource use.

A Health Foundation-supported project at University Hospitals Bristol and Weston NHS Foundation Trust looks to harness data held across multiple, separate systems, using a digital tool and analytics platform to consolidate data into an ICU reporting app. This gives staff the information they need to make key decisions about care, to improve both patient flow and patient safety – helping to save staff time and reduce delayed discharge and readmission rates.

The initial focus is on supporting ICU admissions for elective care patients. The approach has enabled staff to better match ICU capacity with demand, for example, through data on staffing and theatre scheduling. Information about when and why theatre cases are cancelled during the week helps planning to improve patient flow. This is important to clear the backlog in elective care. In the future, the app has the potential to be expanded to support emergency admissions and will become a real-time decision tool.

2. What is preventing this potential from being realised and what can be done about it?

Despite the clear potential for the NHS and its patients to benefit from routine health data and data-driven innovation, there are longstanding barriers that prevent the NHS from realising this potential. Informed by our experience using routine NHS data, supporting others to develop innovations in data analytics, and discussions with others working on these problems, we describe five of these barriers in turn and set out some approaches that might help to overcome them.

Barrier 1: A lack of linked data with detailed clinical information and issues with data quality

Routine NHS data are generated through every patient interaction with the system, and through wider administrative and clinical tasks such as diagnostic tests. These data are collected through a huge number of different systems, which leads to a fragmented data landscape[2] with limited linking of data (as illustrated in Figure 1).

Linking data[3] across services is crucial, allowing health care professionals to access all relevant information about each patient. Patients and carers also benefit from easy access to data held about them and their health. Finally, data linkage is needed for service improvement, research and innovation, and for commissioners and providers to get a complete picture of the volume and quality of care being delivered.

Fragmentation of data remains an issue across the whole of the NHS in England as illustrated in Figure 1 (interactive graphic, best viewed online).

While the data about secondary and community health brought together by NHS Digital through the Secondary Uses Service can be very useful, the limitations of these datasets (which include the Hospital Episode Statistics database, Emergency Care Data Set and Community Services Data Set) have been acknowledged. They lack, for example, detailed clinical information, and have issues with data completeness, driven by their primary purpose being payment and activity monitoring.

---

[2] The NHS's data systems and the datasets generated – and the relationships and flows between these.

[3] Data linkage is the process of joining together records that pertain to the same entity, such as a person or business. (ONS, 2021)

There are increasing efforts to link data at a population level, for example:

- Good progress has been made in joining up primary care data. Local level examples include the Lambeth DataNet, and national level examples include the Clinical Practice Research Datalink (which covers a sample of up to 16 million currently registered patients) and OpenSAFELY (which provides coverage of up to 95% of the population).

- Data across different health and social care services are also being brought together. In North West London, data for 2.3 million patients have been joined up to support direct care (the Whole Systems Integrated Care database), and research and innovation (the Discover dataset). National programmes such as the Local Health and Care Record Exemplar programme have sought to drive progress, though with varying success. The Health Foundation's Networked Data Lab is pioneering the use of linked datasets using federated analytics[4] to help tackle major challenges in health.

Successfully linking data also means addressing legal, ethical and regulatory challenges, which can be resource intensive. For example, bringing together data from primary care can require data-sharing agreements to be put in place with each GP practice. This effort is often repeated to use the data for different purposes. Such measures are necessary to protect privacy, and (along with technical solutions to protecting privacy) are important for ensuring (and demonstrating) the trustworthiness of data use, but could be made easier to navigate to enable faster progress.

Even if joined up, the data collected and held by the NHS give only a partial picture of a patient's health and wellbeing. Reasons for this include:

- Routine data does not include those who do not access health services. Unequal access to health services means the experience of some groups – for example patients from some minority ethnic backgrounds – are less well represented in datasets relative to the level of need that they have. Those missing may be those who have the greatest health need.

- Routine NHS data lacks information about a patient's health in between interactions with the health service. This information could help better understand drivers of ill-health and enable more preventative care.

- Data quality is important, and is often an issue. Data quality can be affected by a number of factors, including choices made by clinicians or clinical coders when entering data,

---

[4] Federated analysis can be described as the practice of developing data science methods which are then applied to local datasets where patient data does not leave the secure local system.

decisions about which data are mandatory to collect (coding of ethnicity data is known to be a problem, for example) and the systems used to record this data.

- The purpose for which data are collected shapes what gets recorded. For example, outpatient administrative data frequently lacks information about diagnosis, because the primary purpose of this data is operational (including ensuring hospitals get paid and managing resources), for which diagnosis is not always required. Even data collected from clinical encounters omits potentially important information about wider patient experience, priorities and symptoms.

Proposed approach: improve underlying infrastructure for data and technology

Strong foundations and good technical infrastructure are essential in order to provide high-quality, timely data for service improvement, research and innovation.

Actions that can help to achieve this, at the national level, include:

- developing technical infrastructure and data standards to address data fragmentation (eg through data linkage and federated analytics) and improve timeliness of data collection and access

- improving data quality and coverage, including addressing sources of bias that might impact health inequalities.

Data collected, and held, beyond the NHS can also be very useful. This 'health-relevant' data includes that held by the wider public sector (such as housing data held by local authorities, and data about employment), and citizen-generated data held by the private sector (eg data from fitness trackers, social media and retail). Exploring these sources could help create a more complete picture of individual and population health and wellbeing, and of the wider determinants of health. There are many challenges to ensure that such data sharing and use is legal, ethical and publicly acceptable, but efforts such as the Open Life Data Framework and work by the Wellcome Trust seek to demonstrate how this can be done. While the most pressing need is to ensure the NHS can make best use of the data generated within the system, it is important that we continue to explore the potential of health-relevant data over the medium-term.

Issues with data are only part of the problem. The NHS also has an under-developed approach to data science and data-driven innovation. Here we consider the range of factors that contribute to this, and explore possible solutions.

Barrier 2: An underutilised analytical workforce

The NHS employs around 10,000 data professionals and analysts, but has struggled to develop and deploy portable data-driven innovations[5] across the NHS. Our previous report (*Untapped potential*) found that analysts lack opportunities for training and professional development, and need better access to the software tools required to derive insights from data and to develop prototypes of data-driven solutions. The NHS also lacks an established culture of open analytics in which solutions developed in one part of the NHS are routinely shared with other parts.

Proposed approach: develop the analytical workforce and wider analytical capability

Developing such capability requires action at multiple levels – and we have previously set out a detailed set of actions for addressing the shortfall in analytical capability. At the national and organisational level, actions include:

- providing the analytical workforce with the skills, recognition, professional status, leadership and support needed to make a difference

- investing in the software tools the analytical workforce needs.

Among the analytical community, it means committing to open analytics and collaboration – to share skills and knowledge, and support the development and scaling of tools across the NHS.

---

[5] Data-driven innovations involve collecting, using and analysing data. This may include using machine learning or other types of artificial intelligence. Portable data-driven innovations are those that can be quickly and easily adopted across different parts of the NHS, for example because the underlying code and software is published under an open licence.

Barrier 3: A historic focus on the wrong problems

Data science and data-driven innovation aren't necessarily focused on the right problems. Too much time and infrastructure is devoted to reactive and routine tasks, such as performance reporting, that could be streamlined and automated to free up capacity to innovate. Our work has highlighted that analytics teams can be siloed, while managers may not understand how data science can help address current or future challenges. In combination with underdeveloped analytical leadership, this means NHS analysts can often be focused on the wrong problems (this applies to innovations developed in industry as well).

Proposed approach: focus on data-driven innovation as a service

Retuning the focus of data science within the NHS will help shift attention to the most relevant problems. This should involve routine deployment of open-source innovations developed in collaboration with end-users – the Royal College of Paediatrics and Child Health's growth charts project has shown how this can be done.

To make this happen, it's important to support NHS managers (at a national and organisational level) to be good commissioners and consumers of data-driven insights. Collaboration could also be improved (at a national and local level) between data professionals and those leading, delivering and transforming health services. This will ensure that data-driven innovations:

- address the needs of end users – whether they be clinicians, senior managers, national policymakers, patients or carers (or a combination)

- address NHS priorities – whether the innovations are developed within the NHS or by industry partners.

Among the analytical community, analytical teams could also take a product-based approach to health data science. Product owners could work with analysts, service planners and clinicians to map problems, and develop and deploy solutions using agile methodologies.

Barrier 4: Challenges in scaling up quickly and well

The NHS faces multiple challenges in rolling out promising data-driven innovations at pace and at scale. When tools are successfully developed, too few are deployed into routine use across the NHS. At a local level, this is in part because analytical platforms aren't connected, or otherwise set-up to allow these tools to be deployed at scale – for example, integrating tools into EHRs is difficult and expensive. For more advanced tools, especially those using artificial intelligence, and with a direct clinical use, there are also challenges including regulation, monitoring and evaluation to support safe adoption at scale. Challenges common to deploying other kinds of innovation in the NHS are also relevant here.

Proposed approach: build better implementation infrastructure

To successfully scale up the best data-driven innovations in the NHS, better implementation infrastructure (both technological and organisational) is essential. This must be accompanied by effective regulation, monitoring and evaluation to ensure safety and equity, and build confidence among health care professionals and the public.

What's needed at a national level is to:

- develop new approaches to actively monitor and evaluate data-driven innovations, to manage any uncertainty and minimise risk

- ensure the regulatory environment is easy to navigate for innovators but mitigates potential harms.

Barrier 5: Risks to safety, outcomes, health inequalities and public trust

There are risks if any drive to increase the use of data and data science in the NHS is not supported by a responsible approach to innovation. Beyond the risks to safety and outcomes that regulation, monitoring and evaluation could address, there are also potential risks to health inequalities, through bias in data-driven innovations, and to public trust if data is misused, or privacy breached.

Proposed approach: foster a responsible innovation approach

To ensure everyone's health care will benefit, a responsible approach to innovation is crucial. To begin with, diversity must be improved across the data and technology workforce at all levels.

Furthermore, at a national and organisational level, we must:

- involve patients and the public in setting expectations and rules for data collection, sharing and use

- ensure transparency in how data are used, and develop approaches to privacy-preserving data science (including through federated analytics and trusted research environments[6])

- provide transparency about who can access data and for what purpose, what the benefits are, and what options patients have

- develop approaches to understand, measure and mitigate possible bias and impact on inequalities as part of in data-driven innovation – at all steps in the process. The Health Foundation is working on this through partnerships with the NHS AI Lab and the Ada Lovelace Institute.

---

[6] A trusted research environment is a secure computing environment that holds data and enables access to it for analysis. (Understanding Patient Data, 2021)

3. Will national policy help realise the potential of data and data science?

The draft data strategy for health and social care sets out the government's ambition to transform how the NHS in England uses data. The strategy represents a welcome step forward, and if successfully delivered, could make a substantial contribution to addressing some of the barriers we have set out in this long read.

While many of the commitments included in the draft strategy are important, they are not straightforward to deliver. In this regard, the final version of the strategy (expected in the coming months) could be strengthened by including a credible delivery plan. This could set out how commitments will be achieved, who will be responsible, what dependencies there are, and crucially, what resources will be made available.

This is not the first time the NHS has set out an ambition to make big strides in the use of data. Previous efforts include the National Programme for IT (which was stood down in 2011 after a report from the Major Projects Authority concluded that 'the National Programme for IT has not and cannot deliver to its original intent') and Personalised Health and Care 2020.

It is important that the lessons identified, for example through the Wachter Review and the National Audit Office's report on digital transformation in the NHS, inform the implementation of the data strategy. Key lessons identified by Robert Wachter included: lack of engagement with clinical staff and over-focus on technology; a centralised approach with insufficient local support; rushed deployment of the programme; lack of central support for NHS trusts; problems with procurement and contracting; changes to leadership and shortages of individuals with the right skills.

There are two other important areas in which the draft strategy could be further developed.

When it comes to health inequalities, the final strategy could demonstrate greater awareness of the risks posed by the increased use of data and data-driven innovations, and the actions needed to address these risks. For example, recognising that due to poorer access to health services some groups may be under-represented in NHS datasets relative to their level of need, and therefore use of this data might not equitably benefit patients from these groups. Lack of diversity in the technology and data workforce is another issue that the data strategy could set out a plan to help address.

The strategy has a focus on improving transparency and building public understanding of how data are used. It could go further here and develop a more comprehensive offer, to involve and engage the public in decisions about the use of data. This, along with technical solutions such as trusted research environments (being considered by the Goldacre review into the use of health data for research and analysis), is an important element of developing a trustworthy approach to using data.

Wider opportunities to achieve change

The NHS data strategy is not the only way to drive change in how the NHS uses data. There are important opportunities to create the conditions needed for change with the forthcoming national strategy for AI in health and adult social care, and the Health and Care Bill (which includes legislation to support the data strategy's ambitions).

There is also potential to align efforts to improve how the NHS uses data with efforts to address NHS priorities through wider transformation programmes. Opportunities here include the elective care transformation programme, and the ongoing transition to integrated care systems. It is programmes like these that will drive changes in the NHS, where improvements in the use of data can make a real difference. Of course, there is risk of confusion over ownership and accountability, of duplicated efforts, and of important issues possibly falling through the cracks.

The recent Wade-Gery review (*Putting data, digital and tech at the heart of transforming the NHS*) recognised the need to align transformation efforts with efforts to make better use of data (as well as digital technology more widely). The bringing together of national responsibility for digital and data in a single 'transformation directorate' within NHS England and NHS Improvement is a good opportunity to deliver such alignment.

The national reorganisation is unlikely to be a universal remedy and questions remain about the extent to which this reorganisation will bridge the gap between innovation and implementation where it really matters – at a local level.

The engagement and hard work of local NHS organisations will be critical to delivering much of the change needed to drive forward the better use of data and ensure it has a positive impact. These local organisations need much more clarity on why better use of data is important, what actions they should take, and where funding and other support will come from.

In conclusion, there is significant untapped potential in data and data-driven innovation within the NHS in England, which, if tapped, could help respond to the unprecedented pressure it faces. However, this full potential will not be realised unless the NHS addresses longstanding barriers that have prevented the development and deployment of responsible data-driven innovation at scale. Achieving this will require a coordinated programme of action, and a policy environment focused on integrating efforts on data with wider transformation efforts.

Supporting information

About the authors

- Josh Keith (@joshkeith) is a Senior Fellow in the Data Analytics team at the Health Foundation
- Fiona Grimm (@fiona_grimm) is a Principal Data Analyst in the Data Analytics team at the Health Foundation
- Adam Steventon (@AdamMSteventon) is Director of Data Analytics at the Health Foundation.

Acknowledgements

This long read was published originally on 26 January 2022 at the following address: www.health.org.uk/publications/long-reads/how-better-use-of-data-can-help-address-key-challenges-facing-the-nhs