HARNESSING DATA SCIENCE AND AI IN HEALTHCARE

FROM POLICY TO PRACTICE

Report of the WISH Data Science and AI Forum 2018

Giles Colclough
Grail Dorling
Farhad Riahi
Saira Ghafur
Aziz Sheikh

مؤسسة قطر
Qatar Foundation

WISH
World Innovation Summit for Health

Suggested reference for this report: Colclough G, Dorling G, Riahi F, Ghafur S, Sheikh A. Harnessing Data Science and AI in Healthcare: From policy to practice. Doha, Qatar: World Innovation Summit for Health, 2018

ISBN: 978-1-912865-01-7

HARNESSING DATA SCIENCE AND AI IN HEALTHCARE
FROM POLICY TO PRACTICE

Report of the WISH Data Science and AI Forum 2018

CONTENTS

FOREWORD

The promise that data science and artificial intelligence (AI) hold to transform the delivery of healthcare is undeniable. Healthcare as a sector, with all of the longitudinal data it holds on patients across their lifetimes, is positioned to take advantage of what data science and AI have to offer. From diagnostics, interpretation of lab tests and scheduling appointments to personalizing care, finding cures to conditions, and creating new and innovative solutions to long-standing problems – the opportunities are endless.

However, compared to other sectors, healthcare lags behind. We have seen how data and technology have helped to transform how these sectors deliver services to consumers; we need to learn from these examples and apply these lessons to the unique context of healthcare.

At the policy level, we see common problems, including the challenges of collection, curation and storage of data. We need to ensure that there are appropriate security and governance measures in place, to tackle the interoperability of software and to make sure that we have trained experts who can work with the growing complexity of data systems.

To maximize the potential of the data assets that health systems hold, we need clear policies to support their use and shape the future. Through our work on this forum, we have had the pleasure of working with a unique group of global experts with backgrounds in academia, policy, clinical care and industry to tackle the key policy issues in this area. We have formulated a series of recommendations that we hope will assist policymakers globally to realize the promise of data science and AI.

Professor the Lord Darzi of Denham, OM, KBE, PC, FRS
Executive Chair, WISH, Qatar Foundation
Director, Institute of Global Health Innovation, Imperial College London

Professor Aziz Sheikh, OBE, FRSE, FMedSci
Professor of Primary Care Research & Development and Director, Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh

EXECUTIVE SUMMARY

Health system leaders across the world share common goals: to prevent, cure and manage illness; to deliver the best possible citizen and patient experience; and to do so in a financially sustainable way.

Over the past century, health system leaders have progressed toward these goals, aided by advances in science and technology: new vaccines, medicines and surgical techniques; technologies, such as telehealthcare, which can dramatically improve access; and analytics to better measure the costs and variations of care provision. These factors contribute to improvements in life expectancy across the globe.

This report by the World Innovation Summit for Health (WISH) has brought together some of the world's leading experts in health policy, data science and healthcare reform. We argue that health system leaders today can steer the next wave of progress in healthcare by harnessing advances in data science to generate valuable insights from the large, complex data sets accruing in health systems. Data science is already transforming other major sectors of human activity – from transportation to life sciences and financial services. In healthcare, these technological advances will make services more accessible, effective and efficient. They will help health practitioners diagnose people earlier, treat them faster and more effectively, and provide new opportunities for patient engagement, empowerment and self-care. In short, they will help achieve the triple aim of healthcare reform. Beyond personalizing and optimizing care delivery, data science also offers the promise of supporting health policy decision-making, better integration of healthcare with other sectors, and substantial time and efficiency savings in undertaking research and driving quality improvement initiatives.

However, healthcare is substantially behind many other industries in implementing data science – in one estimate, for example, the US has gained no more than 10–20 percent of the total opportunity.[1]

Health systems will require five features to be successful:

1. Organization-wide data repositories
2. Data governance and security
3. Interoperability of data within and across health systems
4. Data science capabilities
5. Use and repeated reuse of data to improve decision-making and care.

There are many challenges to realizing the full potential of data science in health systems. However, the possible rewards are enormous, and not just for individual patients. It is estimated that healthier and more productive populations would increase economic output by $10 trillion globally.[2]

Governments and policymakers can take four actions to start the journey toward data-enabled systems:

1. Provide **national leadership** that includes: ensuring clear executive accountability to liaise with the public and engage with critical stakeholders; developing and disseminating the overarching vision of a data-enabled transformation of health; and providing an accompanying, proportionate, regulatory framework

2. Identify and gain **'quick win' opportunities** (first 12 months) to build engagement, experience and momentum

3. Develop and implement a **medium-term strategy** (over one to three years) to digitize and integrate data sets, establish governance arrangements and create centers of excellence for data science in health

4. Pursue a **longer-term transformation plan** (three to 10 years) to embed capability building within education structures and to continually refine data integration and regulatory frameworks.

This work will not be without its difficulties, but the prize for making progress will be substantial for patients. It will also be beneficial for nation states in terms of improved healthcare decision-making and health outcomes, reduced health expenditure, and job creation in data science and research. Given the current global shortage of talent in data science and AI, those countries that delay in embarking on the journey may find it very difficult to make up lost ground.

SECTION 1. DATA SCIENCE AND AI ARE THE TOOLS OF THE 21ST CENTURY

What do 'data science' and 'AI' mean?

Data science refers to drawing insights from large and complex data sets. This includes collating, processing, analyzing and understanding the data. The term 'data science' covers the methods, processes and systems used to do this. It encompasses both traditional statistical methods, applied to far larger and more complex datasets than was thought possible, and newer approaches made possible by artificial intelligence (AI).

AI, a subset of data science, refers to computers that can learn from data and interact with the human world. The goal is to give machines human-like cognition, meaning that they can 'think', and recommend actions based on that thinking, that they can predict outcomes and that they can learn. Today, AI is primarily used to augment human decision-making. However, there is an emerging body of evidence that computers can – in specific, constrained tasks – deliver comparable or even better-than-human performance, usually at greater speed and substantially lower cost.[3,4]

AI can be characterized by four broad technologies:[5]

1. **Natural language processing (NLP):** Siri and Alexa, Twitter analytics and virtual assistants all perform NLP. Technology's ability to analyze human language, extract meaning and sentiment, and reply intelligibly is transforming our communication with each other and with machines

2. **Computer vision:** At the heart of every self-driving car or number-plate recognition system, computer vision comprises the extraction of information from images

3. **Machine learning:** This comprises programs and tools that recognize patterns in data and make predictions based on those patterns – or that can learn from data. The distinctive feature is that the system must learn the mapping from input to output (for example, from a collection of test results to a diagnosis) by itself – it is not provided with an explicit preexisting model

4. **Robotics:** This covers machines' physical navigation and interactions with the human world.

These technologies are not mutually exclusive and are increasingly found within the same AI tools and products.

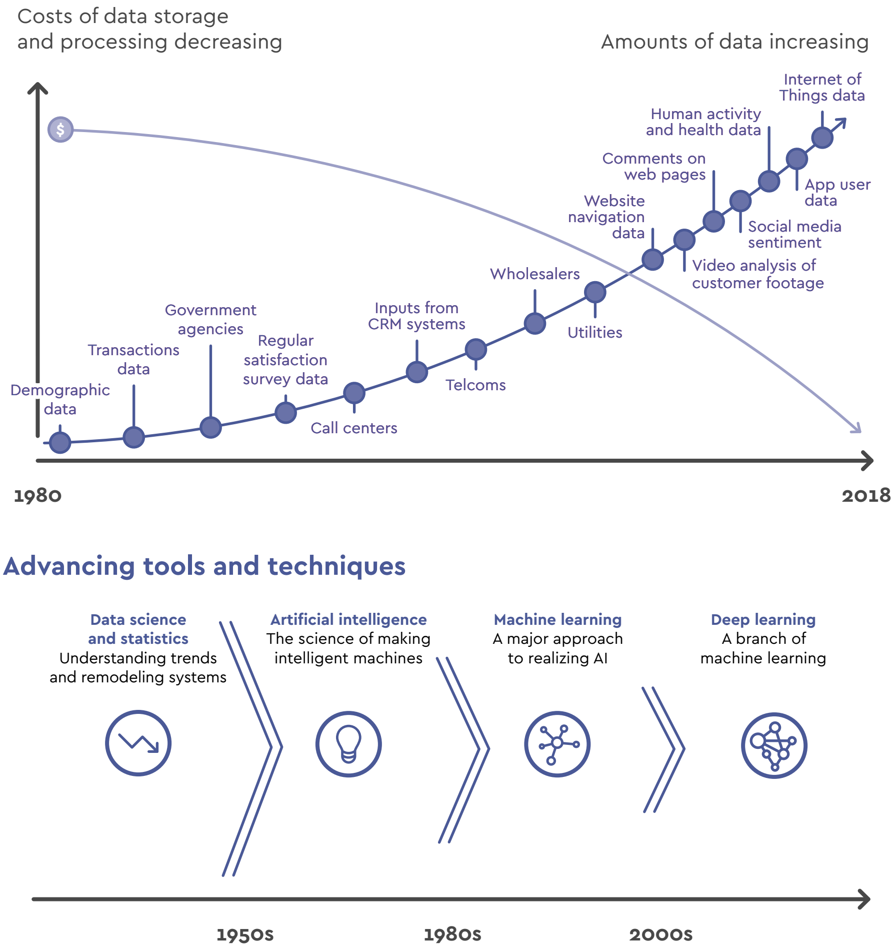
Why is now the time to invest in data science in health?

Over the last 10 years, there have been three shifts in the use of data (see Figure 1). First, people and systems now create much more digital data than ever before, so there is more available for analysis[6] – 90 percent of the world's data was created in the last two years alone.[7] Second, it is now possible to store and process huge data sets more cheaply – at around 10 percent of the cost of a decade ago[8] – thanks to cloud computing and advances in silicon chip technologies. Third, thanks to innovations in the design and use of processing power, complex machine learning can now interpret the data that is created, stored and processed. Together, these three shifts have created an explosion in interest and investment in data science which, while not without risks, has the potential to transform every part of daily life.

Such a transformation is needed more than ever in health systems, which are facing increasingly complicated challenges. More patients are suffering from complex conditions with multiple causes and comorbidities, such as diabetes and child obesity. In more and more cases, even diagnosis requires input from several specialties. This complexity cannot be easily modeled with existing tools. However, new data sets and processing technology offer new capabilities for understanding this complexity. There is further potential benefit in sharing anonymized data with citizens, universities and industry, in open data formats, to develop shared intelligence that emerges from collaboration, collective effort and competition. (For example, the Opendata Transparency area of the Portuguese NHS uses application programing interfaces for data interoperability.)[9] This convergence of need and opportunity makes now the ideal time to invest in data science in health systems.[10]

Consumer enthusiasm to participate in this technology is substantial across the world. Already, 35–40 percent of people in India and Qatar with internet access have access to their health records through a computer or mobile device. This proportion rises above 45 percent in the US and China.[11] For people with an electronic health record (EHR), more than 50 percent are active users (that is, downloaded their EHR in the last three months).[12] In regions where patients have less reported access to EHRs, such as the UK (at 14 percent), almost 60 percent would find such access useful. Consistently across these five regions (the US, China, India, Qatar and the UK), 30–60 percent of people report that they would want their health data to be shared to improve care delivery, to conduct research and to inform health planning. (Other surveys found that more than 75 percent of Americans are interested in sharing their health information to get better care, and about 60 percent of British people would allow their health data to be used for medical research.)[13–15] This consumer interest also extends to AI, with 55–65 percent of people stating that they would be happy for their clinicians to use AI decision-support software in their care – although there is much lower approval for the use of autonomous AI clinical decision-making systems.[16]

Figure 1. Data availability, storage and analytics approaches, 1990–2018



Source: Dave Evans, *The Internet of Things: How the next evolution of the internet is changing everything* (2011)

SECTION 2. DATA SCIENCE CAN HELP TRANSFORM HEALTHCARE

Many health systems are facing the same challenges: predicting and preventing the onset of avoidable disease; determining the safest and most effective treatment option; and delivering cost-effective care. Data science can help address all three challenges. From risk detection, triage and scheduling, through diagnosis, prescription and treatment, to patient engagement and public health, data science can contribute to efforts to optimize resources, better engage patients and improve the quality and outcomes of care.

However, healthcare has lagged behind other sectors in delivering on this potential. AI adoption in healthcare is low, and the projected increases in spending are well below other leading sectors (see Figure 2). Research by the McKinsey Global Institute found that the US has gained no more than 10–20 percent of the opportunity offered by data science and AI in health-care*, compared to 30–60 percent in retail and location-based services.[17]

Specifically, healthcare is behind in delivering three of the building blocks needed for successful data science and AI integration:

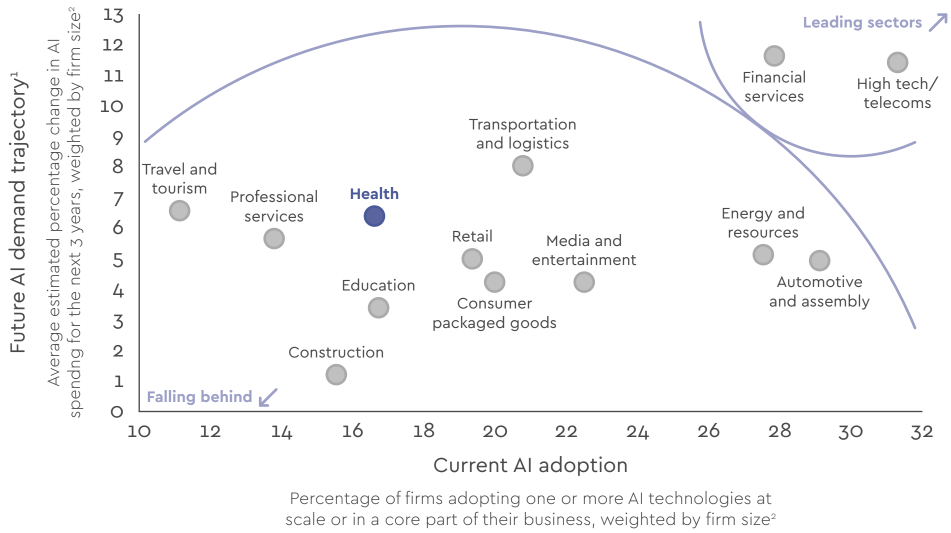
1. Creating the large, integrated and interoperable data sets necessary for the development of new tools

2. Establishing the governance structures and security measures required to handle sensitive data – although some countries are making some headway with this, such as the Health Insurance Portability and Accountability Act (HIPAA) of 1996 in the US

3. Attracting the scientific talent needed to design complex systems. This delay is likely due, in part, to the relatively fragmented nature of the healthcare sector and the economic realities of public healthcare systems. However, the experience of other sectors shows that these challenges can be over-come with sufficient leadership and commitment from policymakers.

* Estimate based on analysis of the proportion of savings identified in 2011 that had been gained by 2015 in the US health sector and the European public sector.

Figure 2. Sectors leading in AI adoption today



¹Based on the midpoint of the range selected by the survey respondent.
²Results are weighted by firm size.
Source: McKinsey Global Institute, *AI adoption and use survey*

How has data science improved performance in other sectors?

Other sectors have used data science to deliver dramatic improvements in output, cost and customer experience. In some cases, these changes have happened much faster than we imagined just five years ago.

- Car automation is a high-profile illustration of data science in action. Today, cars are available that can brake or swerve to avoid collisions, and park by themselves. These systems use computer vision and machine learning algorithms. This kind of automation frees up drivers' time, and while not infallible, is predicted to lead to substantial improvements in road safety. This same approach can be applied to many complex healthcare processes. Some new hospitals, such as Humber River in Toronto, have already automated medication dispensing and internal supply delivery.[18]

- Clinical research organizations use data science to identify the optimal mix of sites in clinical trials. They apply predictive risk algorithms incorporating NLP and machine learning to select sites that are most likely to recruit eligible participants and to meet trial milestones on time and with appropriate data quality. This has increased enrollment time by 15 percent, reduced costs of patient visits by 10 percent and improved quality of targeting by 40 percent.[19] Similar improvements can be expected for comparable areas of healthcare research, meaning that more research could be done with limited public funds.

- Many parts of China have moved to a paperless and cashless society, with the near-complete engagement of consumers on the WeChat platform. A Beijing resident could start her day by checking out a friend's new haircut on WeChat, then booking an appointment at the same salon. From the salon chair she can make a reservation for lunch at a local restaurant, and preorder and prepay for the food. The platform even allows for donations to rough sleepers by scanning a Quick Response (QR) code. All of this, without ever leaving the WeChat app.

What benefits could data science bring to health systems?

If similar techniques were implemented in health systems, it would be possible to achieve the following:

- **Better forecast population trends**: Data science can help to more accurately predict disease burden and costs, identify high-risk patient groups and target prevention therapies. Johnson & Johnson already use machine learning to predict demand for their pharmaceuticals. Mature health systems could save up to 10 percent of total care costs by forecasting population health and targeting patient screening.[20]

- **Deliver more preventative care:** Most healthcare expenditure today is focused on treatment rather than prevention (only around 9 percent of total US health expenditure is related to prevention).[21] Data science can expand research into risk factors and biomarkers, support earlier treatment and, ultimately, enable interventions that prevent disease from occurring.[22] Screening programs could become more targeted and accurate, reducing the risks and costs associated with false positive identification.

- **Further personalize treatments:** Using results from large population studies (such as the Qatar and UK Biobank projects), AI could combine genetics, biology, behavior and patient preference to select the most effective and appropriate treatment pathways. Personalized treatment could be more effective and feel more human than generic models of care. It is estimated that these changes could increase average life expectancy by up to 1.3 years and deliver a global economic impact of $2–10 trillion.[23]

- **Improve user experience:** Currently, it is estimated that US doctors only draw on about 20 percent of available trial results when diagnosing and treating cancer patients.[24] In the future, virtual assistants will provide clinicians with the latest research at a single voice command. Patients, too, will have access to health advice and lifestyle support from their mobile phones. Both Babylon Health in the UK and Gyant in the US already run an AI-enabled chatbot for triage and can support live video consultations with primary care doctors.[25,26]

- **Improve productivity and reduce costs:** AI can automate and optimize tasks across a hospital. AI-supported diagnostic tools could be faster, cheaper and more accurate than current practice. Triage and scheduling could be supported by efficient algorithms. In nursing alone, 30–50 percent improvements in productivity may be possible using AI-enabled tools.[27] System-wide efficiencies, including the automation of repetitive tasks, could lower health spending as a share of gross domestic product (GDP) by two base points in developed economies.[28]

Figures 3 and 4 illustrate the experience of a cancer patient in a hospital environment with these types of data science systems.

Figure 3. A data-enabled healthcare ecosystem

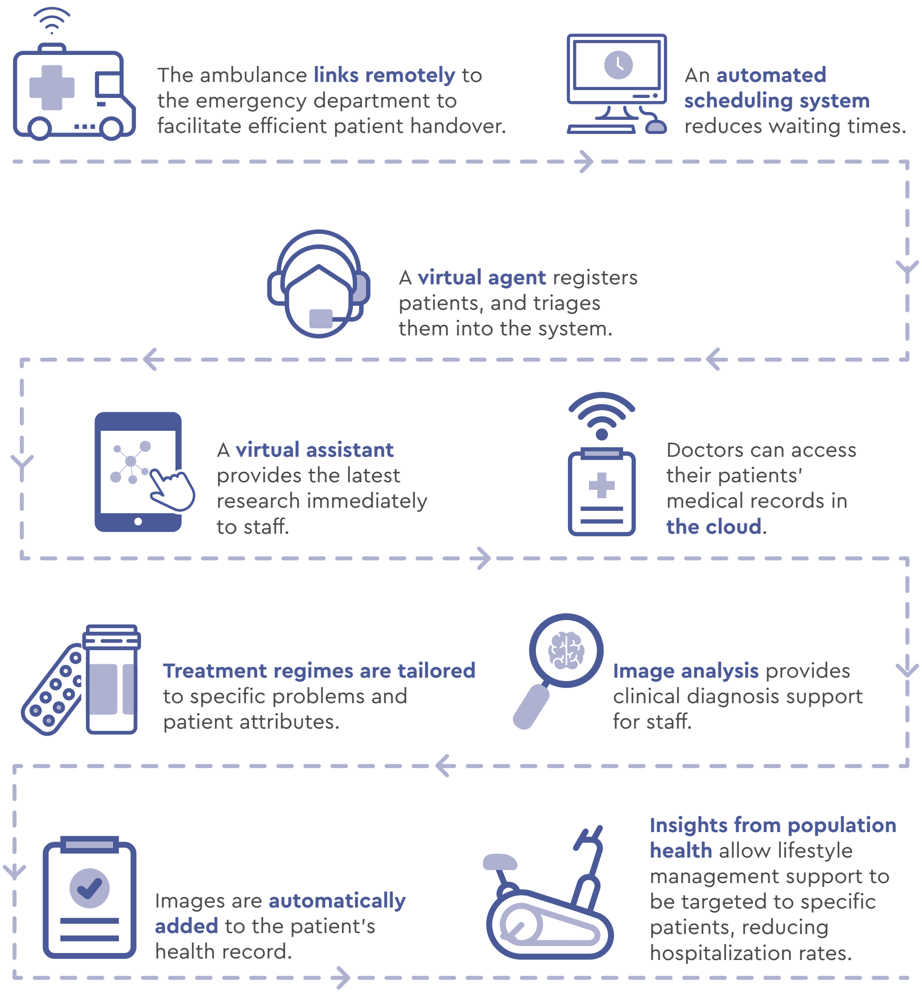
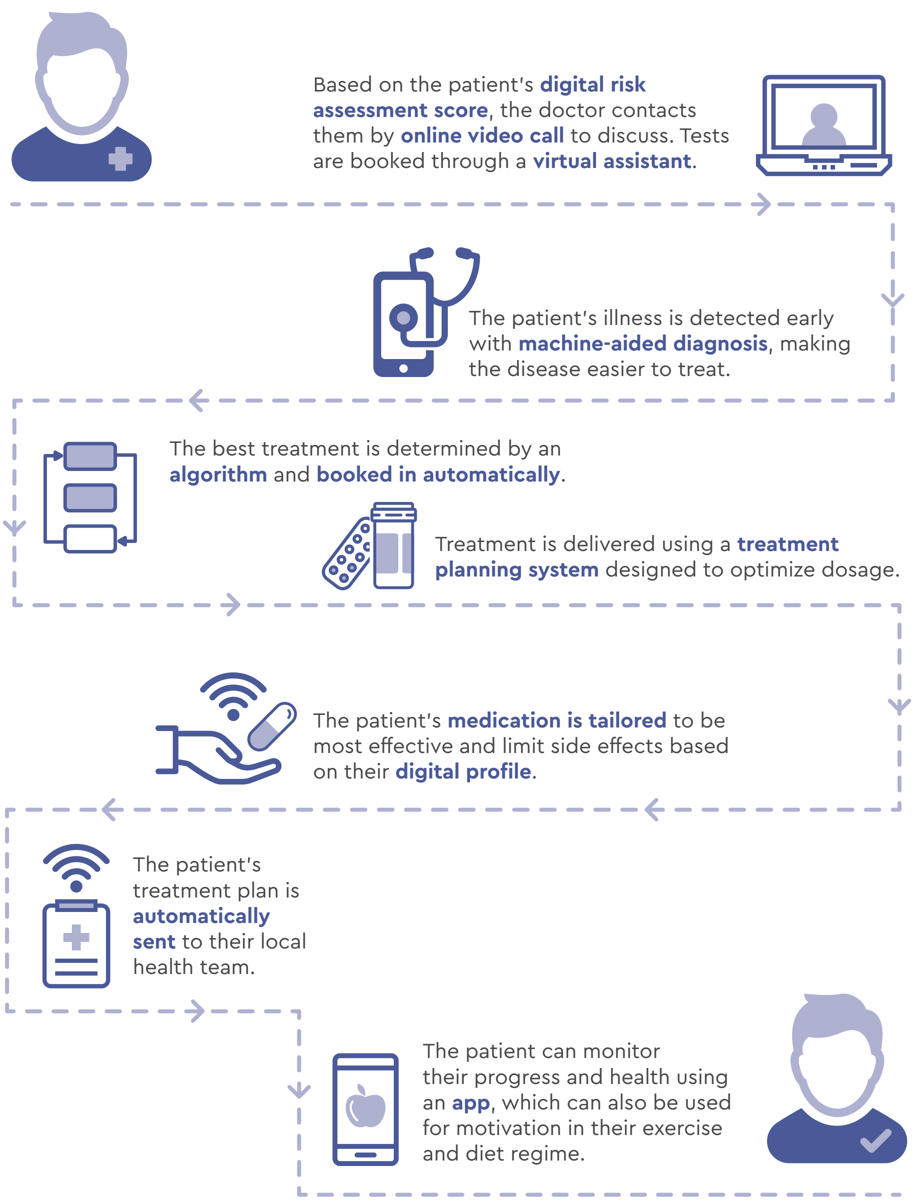


Figure 4. Precision medicine in practice



SECTION 3. MOVING TO A FULLY DATA-ENABLED LEARNING HEALTH SYSTEM

To date, no sector has made the transition to a fully data-enabled system. However, lessons can be learnt from health systems that have taken initial steps, as well as from other, more advanced sectors:

- Interim steps in themselves provide major benefits; it is not 'all or nothing'
- Different parts of the system can progress at different speeds, and early adopters can inspire others
- Difficult questions, such as how to manage data security, have already been answered elsewhere (for example, in the financial sector) and can be adapted for healthcare
- The need for new skill sets and roles will evolve gradually, meaning there is time to recruit and adapt
- While many of the initial benefits of this journey may be outside the health system itself (for example, in the life sciences sector, academia and technology start-ups), if health systems make the investment, there are effective ways to channel the benefits back into health.

Making a transformation of this scale successful requires a clear vision, sustained leadership and a plan that balances short-term goals with longer-term investment. Without such a plan, there is a risk of making investments that do not bring the right benefits and returns, as was the case for England's National Programme for Information Technology.[29,30]

Table 1 gives an illustration of the transformation journey. Although this shows the journey as a simplified linear process, our experience suggests that some parts of the health system will progress faster than others. The 'early adopters' can demonstrate capabilities and impact, share learning and inspire those in other parts of the system.

Table 1. The route toward data-enabled health systems

	Level of data enablement	Capabilities				
		Clinical care	Self-care	Public health	Research and development	Management and operations
Where we are today	Health records combine paper and digital inputs. New data sets collected to answer each question.	Decisions rely on clinicians' knowledge and judgement.	Limited use of apps and wearables for self-management.	Services/policies targeted at high-level risk groups.	Small clinical trials based on bespoke data sets. Some wider data collaborations in imaging and genomics.	Planning based on historic data and modelled forecasts.
Integrated interoperable electronic health records	EHRs used across settings and accessible to the clinician at the point of care.	Clinicians make evidence-guided decisions based on a complete view of patients' data.	Patients can view and contribute to their own records.	Refined targeting of services based on a comprehensive view of risk factors.	Research and patient recruitment is accelerated by larger integrated data sets.	Automatic linking of records and test results improves system efficiency.
Routine use of algorithms for decision support	Whole system data sets analysed offline, and some online prediction systems used in diagnostics, point of care devices and ICUs.	Clinicians' decisions improved by diagnostic algorithms. AI makes low-risk decisions as part of products with regulatory approval.	Beginnings of virtual health assistants.	Accurate, rapid forecasting of infectious disease and population health trends.	Stand-alone diagnostics pass regulatory approval and are incorporated into clinical practice.	Triage and scheduling optimally managed by online algorithms.
Personalized treatment and precision medicine	Genomic and other personal data (wearables, patient apps) integrated with EHRs to make personal health records.	Clinicians deliver precision medicine targeted to the individual's genotype and phenotype supported by AI.	Patients use apps and wearables to monitor and manage wellness and treatment.	Services targeted to individuals based on personal risk score.	Biobank studies yield specific genotypes and phenotypes as predictors of disease.	Accurate scheduling possible by precise prediction of treatment times for each patient.
Integration with broader system and environmental data	Health data sets combined with wider societal data sets, eg: meteorology, pollution, travel, search engine activity, employment, socioeconomic data.	Clinicians rely on a richer view of personal and environmental data, and most decisions are supported by AI.	Advanced virtual health assistants manage wellbeing, and triage into primary care.	Public health plans transformed by the ability to predict disease burden from multiple societal inputs.	Studies modeling effects across whole populations, not patient samples, become routine.	Capacity crises are prevented by accurate models of external system pressures and patient demand.
Fully data-enabled learning health systems	Integrated real-time and real-world data sets and AI are incorporated into all elements of planning and decision-making.	AI learns and improves online, and is used to assess risk and plan optimal treatment in most situations.	Portable drug delivery systems with continuous monitoring liberate patients from hospital beds.	Services are rebalanced to focus on risk management and prevention.	Research outcomes are integrated into health systems using whole population data sets.	System predicts demand and allocates resources preemptively – a system-wide shift to prevention and wellness.

Where most health systems are today

Most health systems today are at the start of the journey. Many high-income countries use digitized EHRs in primary care, but still use a mix of paper records and notes with some digital inputs in secondary care.* Where EHRs are used: they are usually available only to the provider that created them; they are incompatible with other settings, organizations and systems; and they do not routinely use machine learning or NLP.

Clinical guidelines face similar challenges. While guidelines are available for many conditions and settings, they are often static (that is, they require periodic reviews from teams of experts). They are not linked to decision-support tools and EHRs, and they are based solely on the outcomes of published clinical trials. Consequently, clinicians predominately rely on their own learning, experience and judgment to make decisions for individual patients.

Integrated, interoperable EHRs

There are examples where leading health systems have successfully introduced integrated and interoperable EHRs. Just as emails and calendars can now sync across devices, a patient's full health record and care history can be accessed by the treating clinician at the point of care in all settings. Not only does this eliminate duplication, it also enables better continuity of care.

Some countries, such as Estonia, have invested in personal health records (PHRs).[31] These can be managed or even owned by patients, and can combine patients' histories, scans and test results with data from wearable technologies and lifestyle apps. Patients can take ownership of their health data and make decisions accordingly. At the same time, the large data set can help improve clinical practice. Similar innovations are taking place in finance, where the UK's 'open banking' standards and regulations are transferring ownership of financial data to consumers themselves.[32]

* The US is somewhat of an exception, where EHRs are more established and prevalent in secondary care settings.

Routine use of algorithms for decision support

Innovation today is focused on using algorithms to make decisions. Complex algorithms are increasingly able to diagnose patients more accurately than clinicians, and such tools are already starting to gain regulatory approval (see Case study 5). Even simple models built on large data sets can transform clinical practice. For example, a model used at US care consortium Kaiser Permanente to predict sepsis risk reduced antibiotic use in infants by 50 percent.[33]

It is not just clinical decisions that can be improved. Algorithms can also transform the efficiency of hospitals and primary care facilities. For example, patients can now receive health advice and efficient triage into primary care services using mobile apps such as Tencent's WeDoctor, Ping An's Good Doctor and Gyant.

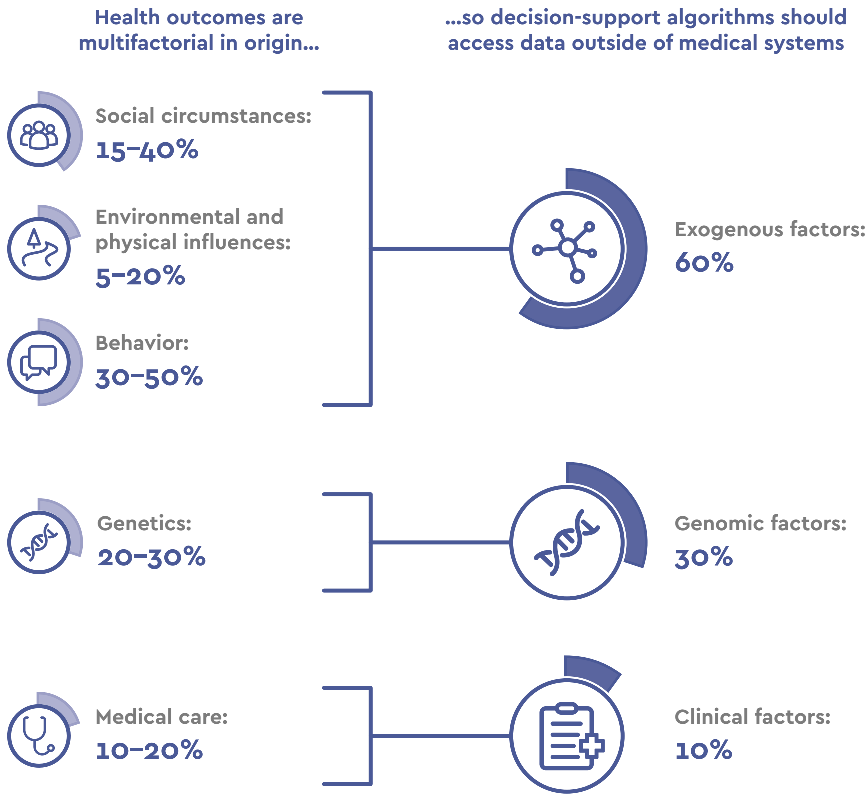
Personalized treatment and precision medicine

Today, most treatment decisions follow standardized guidelines based on clinical trials. However, participants in clinical trials tend to differ from treatment populations in the real world, which can limit the predictive power of the published evidence. Flatiron Health are gathering and analyzing real-world data from EHRs to augment evidence for new therapies.[34]

In a few disease areas, such as cystic fibrosis or some cancers,[35] treatment can now be precisely tailored to the genetics of the patient or tumor. For example, at the Dana-Farber Cancer Institute, clinicians routinely use genomic sequencing in 40 percent of leukemia and lung cancer patients to select specific, targeted therapies. Recent results demonstrate that genetics can also be used to identify patients at risk of steroid-induced growth stunting.[36]

These early steps show promise, but, the potential of personalized medicine is much greater. Health outcomes are influenced by a wide range of factors, including genetics, behaviors and social and environmental circumstances (see Figure 5). Data science could consider all these factors using data inputs such as EHRs, images and wearable devices to create a personal treatment plan. This will help identify risk factors for diseases and biomarkers for effective treatment regimes. Organizations are already investing in acquiring and analyzing such data for population-size samples of patients (for example, the UK Biobank, deCODE genetics, CARTaGENE biobank, Qatar Biobank for Medical Research, Estonian Genome Project and the Nord-Trøndelag Health Study).

Figure 5. Factors affecting health outcomes



Source: "Health policy brief: The relative contribution of multiple determinants to health outcomes", *Health Affairs* (2014)

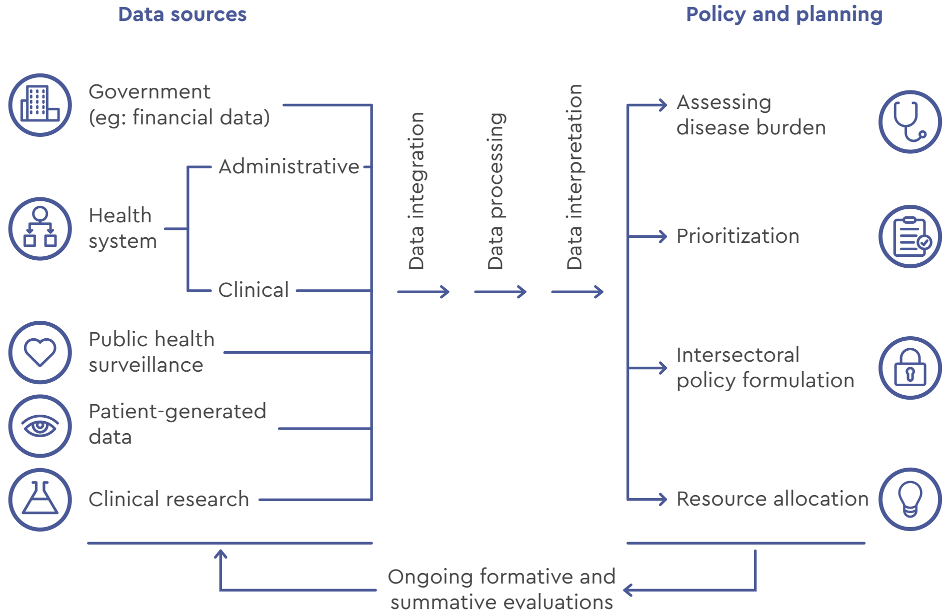
Integration with broader system* and environmental data

Healthcare's finite resources should be allocated according to data and evidence. Currently, this is something that is not always done well. Visualization systems that show resources' capacity and how they are performing *can* deliver great benefits. In Portugal, a map showing hospitals' monthly waiting times was made public, allowing general practitioners (GPs) and their patients to make informed choices about referrals. This was so successful at optimizing utilization that the government lowered their maximum outpatient waiting time guarantees by 30–90 days for some services.[37] The Portuguese national health system also launched the free MySNS Tempos app, which shows users real-time waiting times for emergency departments to encourage a more rational use of services.[38–40]

* For more information, see: Katikireddi SV et al. "Assessment of health care, hospital admissions, and mortality by ethnicity: Population-based cohort study of health-system performance in Scotland", *The Lancet Public Health* (2018).

However, there is an even greater potential in the integration of health data sets with those that include education, transportation, pollution and other societal data. Together, these data sets will enable more precise forecasting and more intelligent use of scarce resources (see Figure 6). One early example is the Artificial Intelligence in Medical Epidemiology (AIME) platform, which predicts the timing, geographic location and spread of a dengue fever outbreak, three months in advance, with 86 percent accuracy.[41]

Figure 6. Data science and policy, planning and evaluation



Fully data-enabled learning health systems

The convergence of all data science technologies – including machine learning, wearables, robotics, networked smart sensors, bioengineering and molecular biology – will comprehensively transform the way care is planned, delivered and experienced. Components of this are already becoming a reality.

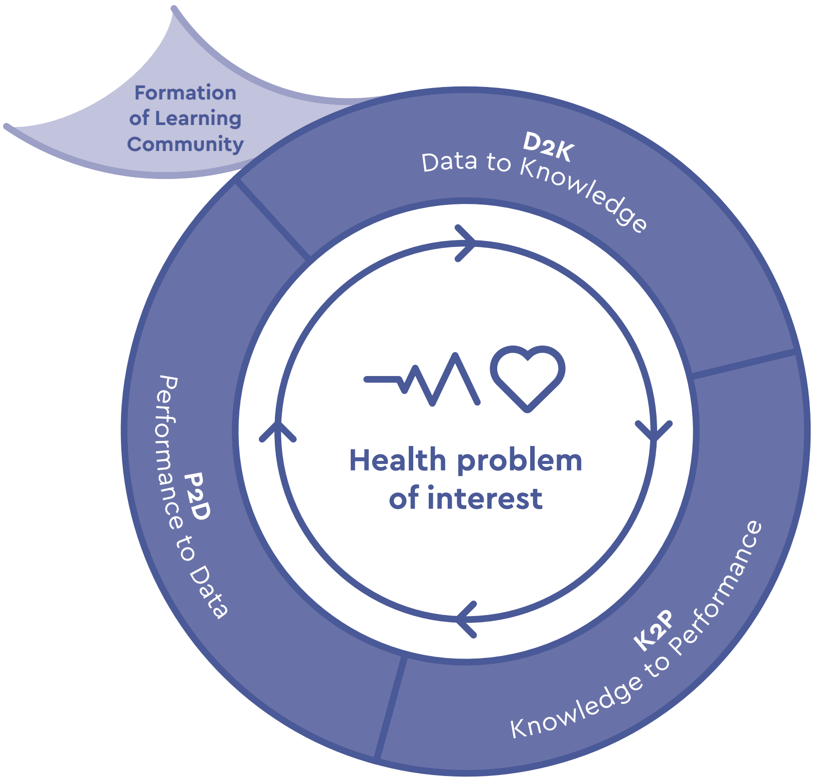
Robot-assisted and robot-delivered services already exist in the most advanced hospitals, including robot-assisted surgery, robotic porters and robotic dispensers. Furthermore, work is underway to create therapeutic systems that continuously optimize treatments or medication dosages while monitoring patient response. One notable example is the creation of an artificial pancreas by Medtronic and IBM Watson.

New 'smart', paperless hospitals are being built around the world with sensor-enabled operating theaters, allowing the automation of many processes and freeing up nearly 50 percent of staff time. As an example, the University of Pittsburgh Medical Center is investing $2 billion in three new hospitals, which are being designed in collaboration with Microsoft and will take advantage of the latest technology and data analysis. An emergency department case study measured a 26 percent reduction in workload following partial automation, and a 48 percent reduction following full automation. A reduction in workload does not necessarily imply a reduction in staff, but can lead to redeployment of staff into more useful activities.[42, 43]

Ultimately, we will see AI-enabled healthcare facilities linking to 'smart homes' and 'smart cities' that will monitor individuals' health behaviors and biometric signs. These will prompt patients and healthcare professionals to make better informed decisions and more timely interventions.

This convergence will allow health systems to continuously learn and improve, on the basis of routine, systematized and automated analysis of the performance data generated. In a learning health system, a cyclical process is set up, whereby data sets are converted into knowledge, knowledge is translated into practice and the results of that implementation create new data, which converts into further insight, which feeds into further stages of the cycle (as shown in Figure 7).[44]

Figure 7. The Learning Health System: From data to knowledge to performance



Source: Adapted from Friedman, CP et al (2017)[45]

SECTION 4. THE FIVE BUILDING BLOCKS OF TRANSFORMATION

To realize the potential of data science in health, policymakers need to create the environment and conditions for it to flourish. This includes creating the necessary infrastructure and incentives, but the starting point is an effective strategy. Such a strategy should include five building blocks (see Figure 8):

1. Organization-wide data repositories
2. Data governance and security
3. Interoperability of data within and across health systems
4. Data science capabilities
5. Use and repeated reuse of data to improve decision-making and care.

1. Organization-wide data repositories

Data science requires high-quality, accessible data sources with large sample sizes. To reap the benefits of data science, organizations first need to gather and store data digitally. For example, a collection of carefully labeled, curated data sets of thousands of skin melanoma images can be used to develop a tool that can diagnose malignant tumors more accurately than human specialists (see Case study 5). To collect the volume of images required, several test centers must use the same systems and, work to the same quality standards and share their results.[46]

Single-application tools like this are of great value, but the potential impact grows as the scale of the data repository and its linkages increase. Collating EHRs with scans, pathology reports, test results and more will enable unprecedented improvements in care, patient outcomes and operational efficiency. As with the skin melanoma images, the key to success is collating detailed information on millions of individuals.

In 2008, Estonia implemented an eRecord system (see Case study 1) that digitally records and stores every interaction a patient has with the health system. This benefits patients because it avoids the need to repeat their medical history on every admission. It also provides population-level data sets that can help in the development of new clinical tools and operational processes.

Figure 8. Building blocks for data science in health

USE/REUSE OF DATA

- Encourage health providers to **work with industry** (tech and life sciences) and **academia**
- Explore and pursue **models of patient consent**
- Support **dedicated multidisciplinary environments** and centers of excellence.

INVESTMENT IN SKILLS AND CAPACITY

- **Support the recruitment** of data scientists, coders and analysts into health from other sectors
- **Create career paths** and **hybrid training programs**
- **Assess the impact on clinical roles** and provide **education and training** to adapt to AI-enabled care delivery
- **Partner** with academia and leading industry players, and **embed experts** within healthcare.

GOVERNANCE AND SECURITY

- **Communicate** the value and purpose of data integration and analytics to build trust and understanding
- Support investment in **cybersecurity** and **validation of technologies**
- **Drive balanced regulations** to protect patients' rights while allowing systems to innovate and improve
- Establish routes of legal liability for AI-delivered services and AI techniques in health eg: natural language processing.

ORGANIZATION-WIDE DATA REPOSITORIES

- Invest in understanding the **value of data** and **improving data quality**
- Encourage or enforce organizations to **adopt EHRs to prescribed specifications** using a range of incentives.

INTEGRATION AND INTEROPERABILITY

- Encourage or enforce **common standards** to support interoperability and **open API (application program interface)**
- **Facilitate the integration** of data sets across organizations, especially in the public sector.

Source: Derived from Cresswell KM et al[47]

2. Data governance and security

Clearly, data repositories that hold sensitive and potentially identifiable patient information are a security concern, and patients understandably worry about the inappropriate sharing and use of their data. Healthcare is particularly vulnerable to cyberattack due to its limited resources, fragmented governance structures and cultural behaviors. Compared with other critical sectors, healthcare has chronically underinvested in IT, and is one of the most targeted sectors for cybercrime globally. The 2017 disruption to the National Health Service (NHS) in the UK caused by WannaCry ransomware was a prime example of a system-wide incident that affected more than 600 organizations and many thousands of patients.[48]

Good security is about controlling risk, reducing vulnerability and mitigating impact. All healthcare organizations need to respond to the growing cyber-security challenge, and policymakers must instigate careful governance and security arrangements, and communicate these arrangements to the public.

Architects of data repositories must answer five questions:

1. Who can access which data and under what circumstances?

2. How do we prevent and detect unauthorized access to data?

3. How do we ensure the accuracy and consistency of data?

4. How do we ensure traceability and accountability for each interaction made with the data?

5. Where does liability rest for breaches of data security or erroneous recommendations arising from automated algorithms?

In designing a system that addresses these points, a few basic principles apply:

- **Only collect, store and give access to essential data:** For example, a company developing systems to improve surgery scheduling may only need access to event timings, diagnoses and outcomes, but not medical histories or scan results. A patient, however, should have access to all data pertaining to themselves. (We note that this provides a tension with AI analysts who typically want access to as much data as possible.)

- **Use unique, consistent but pseudonymous patient identifiers** to link records across the system.[49]

- **Consider using local or cloud-based repositories** rather than one central data bank. This has two advantages, first, it enables access to the original source of data prevents the risk of introducing errors during transmission and replication; second, in the event of a security breach, only a small portion of data would be at risk, not the entire data set.

- **Use encryption and enterprise-grade security measures:** These systems are standardized (for example, in ISO 18033 for IT security techniques) across the financial, telecommunications and consumer industries.

- **Ensure traceability and transparency of all interactions with data:** Estonia allows patients to access a log of all interactions with their data (see Case study 1). DeepMind Health are developing similar processes based on ideas from Blockchain (the Verifiable Data Audit project[50]), to allow hospitals to check that records are being used appropriately.

3. Interoperability of data within and across health systems

AI, and in particular machine-learning techniques, is most powerful when it uses different data types from a range of sources. For example, flu-forecasting tools can combine geolocation data, past flu trends and Twitter feeds to forecast influenza outbreaks with up to six weeks' lead time.[51] Further studies have shown that disease outbreaks can be observed through Google search terms – for example, observing trending search terms that correlate with outbreaks of thunderstorm-induced asthma.[52] For these purposes, data can be integrated at two scales: within an individual tool or device, and at a larger system scale.

Within a single tool or device, the integration of multiple different data sources, even if collected locally, can give powerful results. For example, the continuous monitoring of Rolls-Royce jet engines uses machine-learning systems to fuse multiple data sources in the engine and its surroundings to warn of defects. This data use could transform the evaluation of prosthetic joints. Implant integrity can be monitored live using wearable sensors, using a combination of functional and movement-related metrics, rather than just x-ray images.

At a larger system scale, data can be integrated across an entire national health system and then combined with other public records. One approach to ensuring interoperability of health records is to select a single EHR system across all providers, as in Qatar (see Case study 3). This approach does, however, risk creating a monopoly vendor. An alternative is to restrict EHR choices to a subset with compatible interfaces. Estonia (see Case study 1) has successfully linked EHRs across the country to a wide range of other government eRecords and external data sets. While data sets are stored locally, they can be accessed remotely and linked using unique identifiers for each citizen. Integration like this requires significant mutual co-operation and policy oversight. While this may be challenging in healthcare, it is not unachievable. Several other industries have successfully managed to standardize complex systems, such as competing mobile telecommunications manufacturers that all use the same communication protocols.

Policymakers have an important role to play in promoting common standards and open source tools, to support the development of publicly accessible resources. Successful examples from healthcare include the Digital Imaging and Communications in Medicine (DICOM) standards[53] used in medical imaging, and the open-source clinical management systems, such as Open Source Clinical Application Resource (OSCAR)[54] used in parts of Canada.

4. Data science capabilities

Most health systems do not have advanced capabilities in data science. In the short term, skill gaps can be filled through partnerships and talent acquisition. In the medium- to longer-term, countries will need to train a broad pool of health data scientists.

Attracting and retaining talented staff

Attracting data science talent to work in health systems can be challenging. In part, this is because it is not feasible to offer salaries comparable to the private sector. However, the health sector offers a different value proposition based on the impact and public benefit of the work. Health systems need to invest in dedicated career tracks or hybrid models of employment to attract and retain top talent.

Health systems will further benefit if medical training programs prepare future health professionals for the importance of big data analysis in clinical decision-making, performance monitoring and financial management.

The most efficient way to use a small number of data scientists in a health system is to establish centers of excellence to drive innovation (see Figure 9). These centers must have access to data, the full range of skill sets required and interactions with healthcare leaders. As data science matures, centers of excellence can be used to demonstrate their value and to train new talent.

Figure 9. Skills and disciplines needed by centers of excellence

The key roles within a center of excellence are:

- **Chief information/technology officer** who the center of excellence reports to
- **Data Engineers** who access, assess, import and clean data, and determine the data management approach, hosting environment, security and toolset
- **Data Scientists** who develop models, evaluate performance and transform model outputs into usable tools
- **Delivery Architects** who assess and sustain impact against the team's objectives, and drive adoption of tools throughout the organization
- **Translators** who define and prioritize problems to be solved, and identify and locate required data.

Centers of excellence have been the favored model to underpin large-scale analytics transformations in the insurance, finance and consumer sectors.

Partnership

Health organizations can capitalize on existing expertise by partnering with external agencies. For example, most teaching hospitals already have relationships with academic institutions. Partnering with technology start-ups or industry leaders – who have extensive data science capabilities, but lack clinical expertise – can be another simple way to grow skills quickly. For example, the Beth Israel Deaconess HealthCare system (Case study 4) has partnered with data scientists at Amazon and Google to deliver 11 AI-enabled system innovations in less than two years.

Training

To build the skilled workforce needed in the longer term, governments and health systems need to consider the entire learning arc, beyond primary and secondary education. For example, massive open online courses (MOOCs) can be used to train a workforce on data security. Also, undergraduate and postgraduate courses on data science can be offered to unskilled workers and health professionals. Clinical and executive leadership courses on data and data science are also required, such as that offered by the UK's NHS Digital Academy (Case study 2).

5. Use and reuse of data to improve decision-making and care

The fifth component that health systems must get right is the regulatory environment. To ensure that systems can continue to learn and improve based on the data collected, they must be able to use and repeatedly reuse the data for multiple purposes.[55] Regulatory issues in data science are related to two areas: consent and device approval mechanisms.

Consent

Traditionally, patients have been asked to give informed approval for their data to be used for a specific reason, for example to develop a new cancer-screening tool. However, this model prevents the patient's data from being reused for a different research purpose, limiting the potential of large, integrated data sets.[56]

A new approach, termed *broad consent*,[57] asks patients to agree to their data being held for a much longer period, and being used for a wider range of applications. Consent is still specific regarding who can use the data and for what purpose – for example, for the improvement of the design and provision of higher-quality, lower-cost care – but more exact projects need not be specified. This model has been taken by major population health studies, such as the UK Biobank[58] and the Qatar Genome Programme,[59] and has recently been incorporated in US regulations on the use of human subjects in research.[60]

Device approval mechanisms

Introducing new software for aiding clinical decisions and operational practices may require regulatory approval. This can involve clinical trials to prove that these tools are at least as effective as existing techniques. Although this process is typically straightforward – several image diagnostic companies have won Food and Drug Administration (FDA) and/or European Conformity (CE) approval for the use of AI in clinical diagnostics (see Case study 5) – it can be slow.

Regulators should think carefully about the burden of proof required for potentially transformative solutions, and the imposed time to market. The FDA, for example, has recently downgraded the requirements for radiological cancer diagnostic tools to Class II, meaning that new products need only demonstrate substantially equivalent performance to a current legally marketed device. In addition, the US 21st Century Cures Act set up an expedited access pathway for FDA clearance to fast-track devices that offer significant advantages over existing alternatives.[61] Such changes improve the time to take products to market. This makes it easier for developers, especially start-ups developing a single product, to gain financial backing.

CASE STUDY 1

Estonia's eHealth system

Strategic building blocks
1. Data repositories; 2. Data governance and security; 3. Interoperability of data; 5. Use and reuse of data

Since the late 1990s, Estonia has been moving its 1.3 million citizens to a digital government model. Under the leadership of a national Chief Information Officer, citizens can now vote, manage their tax, manage their education and public services, and control their health records online. The benefits to the citizen are clear: transparency, speed and ease of use. For the state, the benefits are enormous: Estonia estimates that it saves 2 percent of its GDP each year in salaries and expenses through digitization.[62]

There are several important features of Estonia's eHealth system:

- **Every citizen owns their own data:** Each person can share and control access to their data.
- **Each record is linked by a unique citizen identifier (ID):** Each citizen can be identified and linked across all public health databases. Citizens also have a cryptographically assured digital signature, which carries the same legal weight as a written signature.
- **'Once only' policy:** Government will ask for a piece of information about a person only once. Furthermore, data sets are stored where they are generated. For example, a dentist's practice holds its own data, as does a hospital and a GP surgery. This practice greatly increases cybersecurity – if one server is compromised, then only limited amounts of data are at risk.
- **Data sets are linked nationwide by the X-road system:** This securely connects databases across the country over the internet. A doctor can call up patients' histories, scans and test results, accessing information directly from the relevant systems.
- **All interactions with records are logged** using a blockchain-style ledger. This allows patients to monitor every interaction that is made with their data. High-profile cases illustrating the transparency this brings include several GPs who looked at the records of the former Prime Minister after a skiing accident – these doctors were proactively identified and lost their license to practice.

- **Consent management:** Estonia is currently adopting Finland's mydata.org system which allows individuals to provide consent for their data to be shared for scientific purposes, and to charge companies for their use in commercial research.

Key lessons: Estonia demonstrates the importance of early and ongoing public engagement, strong and pragmatic leadership, and clear legal governance. While the country is small, this can be a model to other systems looking to enable a data-science transformation in healthcare.

CASE STUDY 2

The UK's commitment to developing health IT capabilities

Strategic building blocks
1. Data repositories; 3. Interoperability of data;
4. Data science capabilities; 5. Use and reuse of data

The UK was the first country in the world to develop a national health IT strategy. However, this strategy largely failed to realize its potential to improve care and health outcomes. Part of this failure has been attributed to focusing too much on implementing EHRs without investing in how the digital data would or could be used to drive transformational change.

The UK Government has learned from this and other international experiences, and has reprioritized digital maturity and clinical and data science capabilities in the NHS. It has created four training organizations: the NHS Digital Academy to train chief clinical information officers and chief information officers in data science skills and leadership, the Farr Institute, Health Data Research UK (HDR UK), and the Alan Turing Institute. These organizations are simultaneously developing academic health data science capabilities on a national scale. For example, 21 universities are working together in six regional Substantive Sites in HDR UK. There is also a very deliberate attempt, catalyzed by the UK Life Sciences Industrial Strategy, to stimulate data-driven innovation and job and wealth creation.

Key lessons: Health technologies such as EHRs are important, but are insufficient on their own. The focus needs to be on maximizing the opportunities and capabilities to securely and repeatedly use data for clinical, operational and academic benefits. Partnerships working across traditionally competitive organizations are essential.

CASE STUDY 3

Qatar's focus on data-enabled healthcare

Strategic building blocks

1. Data repositories; 3. Interoperability of data;
5. Use and reuse of data

Qatar has implemented one EHR system across the country, making the entire population's health records interoperable. Furthermore, government funding is being used to improve and prioritize patients' interactions with the care system. For example, the Qatar Computing Research Institute (QCRI) are developing AI tools that can translate into different languages for international patients, with around 90 percent accuracy. They are also developing mobile apps that can estimate body mass index (BMI) from facial images, and monitor children's life-style, food habits and sleeping patterns using wearable sensors.[63, 64]

Key lessons: Moving to one single EHR system, or mandating that different EHR systems are compatible, makes it easier to build the population-level data sets needed for AI.

CASE STUDY 4

Beth Israel Deaconess HealthCare system

Strategic building blocks
4. Data science capabilities; 5. Use and reuse of data

The Beth Israel Deaconess Healthcare System includes 2,600 physicians across four hospitals. It is one of the leading centers of data science innovation in US healthcare. Its success has been driven, in part, by mandating tight interoperability of EHRs, and by moving its data hosting to the cloud. However, crucially, the hospital network has also included skilled data engineers in its organization structure.

Amazon and Google employees were brought in and given access to 7 petabytes of data. The strength of this collaboration has greatly increased innovation. Within two years, operational tools were developed to:

- Automate repetitive tasks – AI can classify 99.9 percent of documents and help with information input[65]
- Improve efficiency in operating theaters by 30 percent – AI can predict exactly how much time to schedule for a particular patient–surgeon combination[66]
- Reduce human error in the detection of breast cancer from lymph node biopsies by 85 percent using a deep-learning image analysis tool.[67]

Key lessons: Partnering with existing technology leaders and embedding talent within health organizations facilitates faster innovation and grows capabilities.

CASE STUDY 5
Image analysis tools for clinical decision support

Strategic building blocks
1. Data repositories; 3. Interoperability of data;
5. Use and reuse of data

AI is now able to process images with a high degree of accuracy, even outperforming humans in some cases. This capability can be used to diagnose patient images from anywhere in the world. There are several examples:

- Israeli start-up Zebra Medical Vision offers image analysis tools to help diagnose osteoporosis, emphysema and brain hemorrhages. Since algorithms are cheap to run, this clinical support can cost as little as $1 per scan.
- Ping An Healthcare and Technology developed world-leading AI-based lung nodule detection systems for CT scans,[68] which are being rolled out across hospitals in China.
- A Stanford University platform uses mobile phone images to detect skin cancers as accurately as a dermatologist. This provides a route for quick, accessible screening for melanoma worldwide.[69]
- QuantX is a computer-aided breast cancer diagnosis platform for magnetic resonance imaging (MRI) scans. QuantX received fast-tracked regulatory approval because the US FDA prioritizes treatments that can offer substantial increases in care quality.
- ET Medical Brain, Alibaba's cloud-based solution, combines data hosting and image diagnostics. It is a leading AI-enabled health solution in China, working in partnership with hospitals to access data and develop tools. One tool detects thyroid cancer from ultrasound images with an 85 percent success rate, better than the human rate of 60–70 percent.[70]

Key lessons: Algorithm-based diagnostics are developing quickly, and often deliver rapid return on investment. Smoothing regulatory pathways may lead to more rapid private investment in this area.

CASE STUDY 6

DeepMind Health[71]

Strategic building blocks
2. Data governance and security; 5. Use and reuse of data

DeepMind Health is a British AI company within the Alphabet group. Their health effort was launched in early 2016 and is currently divided between direct-care tools and AI research.

Their direct-care tool, Streams, is used by clinicians at the Royal Free London NHS Foundation Trust. The Streams app escalates the right information about a patient to the right clinician at the right time, reducing the use of paper-based systems, pagers and desktop computers.

Early in the Streams project, the UK Information Commissioner (a data regulator) raised concerns about the Royal Free Trust sharing patient records with DeepMind. In response, DeepMind took steps to become one of the most transparent companies working in health. It appointed a panel of independent reviewers, worked with a group of patients and carers, and published its contracts with NHS partners. DeepMind has now signed partnerships with three further UK-based hospital groups to deploy Streams over the course of 2018.

Key lessons: Giving private companies access to patient data creates obvious security concerns. Careful regulation is required and good practice by private companies can mitigate risk and allay public concerns.

CASE STUDY 7
Flatiron Health

Strategic building block
5. Use and reuse of data

Flatiron Health has designed a technology platform that enables cancer researchers and care providers to use data and analytics to better understand the journey and clinical outcomes of a cancer patient. Care providers enter data which can then be used by other care providers and by pharmaceutical companies for a fee.

Flatiron Health carefully curates its data and extracts information that is often lost in free-text fields. This gives users a much more powerful data set to accelerate research and generate clinical evidence in order to improve patient care.

Key lessons: Allowing the reuse of high-quality data sets between different groups, such as care providers and pharmaceutical companies, for different purposes can maximize the benefits to patients and to the healthcare system as a whole.

SECTION 5. POLICYMAKERS CAN ACT NOW TO START THE JOURNEY

Governments and policymakers who wish to start embedding data science and AI technology into their health systems need to consider taking four actions:

1. Provide national leadership for data science and AI in healthcare

2. Identify and gain 'quick win' opportunities (first 12 months)

3. Set strategic priorities for the medium term (one to three years)

4. Pursue a longer-term transformation plan (three to 10 years).

1. Provide national leadership for data science and AI in healthcare

Responsibility for transitioning to a data-enabled healthcare system could sit either within an individual ministry (such as the Ministry of Health) or be assigned to a cross-governmental entity set up specifically for this purpose. For example, Thailand set up its Health Promotion Foundation, Thai Health, to be chaired by the Prime Minister, with representatives from all relevant ministries on its Board. Countries that have made the most rapid progress in data science – for example, Estonia – tend to have taken this cross-sectoral approach.

The leadership body's responsibilities should include:

- **Communication:** Tell the public about the potential risks of the program, such as security breaches, and how these are being mitigated, as well as the benefits, including greater convenience, more transparency and better patient care

- **Engagement:** Consult and partner with critical stakeholders, especially within the private sector, which is likely to be a major source of investment and innovation

- **Strategic direction-setting:** Develop and execute a strategy for data science and AI that:

    - Sets the regulatory framework governing use of personal data in health (and other settings)

    - Defines the investment budget and plan, the size and scale of which will depend on the conditions of the local health system and model of government (there are other permutations of government)

    - Is consistent with existing government strategies relating to health and digitization.

- **An advisory board:** Set up a national policy group for data and informatics to provide guidance and advice. This board would have technical and academic expertise covering: data set creation, data value, information governance, data security, data integration, capability building, and enabling the use and reuse of data for multiple ends. This board could also include patient and clinician representatives to ensure that the strategy is shaped by the priorities and concerns of these critical constituencies.

2. Identify and gain 'quick win' opportunities (first 12 months)

Ideally, one team within the policy group would focus on finding opportunities for showing an impact in the first 12 months. These 'quick wins' could build on existing data systems and processes or borrow from other countries' successes. The team could create a portfolio of high-value, high-feasibility projects by assessing the ease, speed and cost of implementing an idea, against its potential impact. These quick wins will bring benefits to the health system quickly, and they will also build positive public engagement and trust for longer-term, more significant changes.

Example quick win opportunities might include:[72]

- Launching apps that give patients on-demand video consultations with primary care providers, or launching chatbots to guide patients to appropriate primary care resources

- Adding AI-supported diagnostics to radiology, especially where these reduce waiting times for the scan and the results

- Using triage algorithms in emergency departments to improve flow, reduce waiting times and keep patients informed.

In many health systems, the single most valuable quick win may be to reform regulations to permit cloud-based data storage. Cloud-based storage can currently offer cheaper, secure data hosting, provided there are checks to ensure that personal data are only accessible to the right users (including the patient). Many systems are already making this transition, including Estonia, Scotland, Switzerland and individual health systems in the US. This shift could have an immediate and dramatic effect on the costs and lead times for data integration for clinical care and research.

3. Set strategic priorities for the medium term (one to three years)

With national leadership in place and stakeholder engagement underway, policymakers can develop plans for the next three years. While initiatives will need to be tailored to the existing level of digital maturity, in all systems the critical areas to get right include:

- **Digitization and integration of data:**
    - Agree and mandate the use of basic data formats and standards (for example, for images, medicines and dosage syntax, and for operational data), mutually compatible EHRs and ePrescribing tools within the healthcare system
    - Develop plans to scale existing data sets to a country- or system-level where possible
    - Set up a strategy group to decide how to link different data sets across systems and regions. First steps could include creating unique patient IDs, or starting to link data sets such as primary and secondary health records with mortality data and social care data.
- **Data governance and stewardship:**
    - Establish more permanent structures to maintain the regulatory framework covering personal data sharing and security. This is a priority for many countries – even three years ago, only 17 percent of the World Health Organization's (WHO) member states had a policy or strategy for the use of big data in health. Steps could include:
        - Appointing a chief information/data security officer and strategy sub-group to manage the protection and security of health data
        - Developing the regulatory infrastructure for information governance and data protection – for example, deciding where and how information should be hosted, who defines access permissions and permissible uses, and how to manage data breaches when they occur

- Enacting legislation permitting patient data to be securely hosted in the cloud

- Developing a comprehensive set of policies to manage consent for data use and reuse and, where possible, move to broad consent models in research and routine clinical care. Policies will need to cover the legal basis for consent and the processes by which organizations record and store an individual's consent status.

- **Capabilities and skills:**

  - Promote centers of excellence (see Figure 9) in several clinical hubs serving large populations, and mandate the use of consistent systems and data formats between the centers

  - Establish innovation hubs with appropriate levels of access to data to catalyze private companies and start-ups co-located around centers of excellence.[73]

4. Pursue a longer-term transformation plan (three to 10 years)

  - The longer-term plan can focus on continually improving data integration, refining regulatory frameworks and building capabilities. Actions could include:

- **Continually improving data integration:**

  - Expand from EHRs to patient-controlled personal health records (PHRs), which give access and ownership of data to patients and combine information from health systems with data from patients themselves

  - Build institutional, regional and international relationships to disseminate knowledge and learn from best practice in the construction, use and maintenance of big health data sets. In the longer term, there is the potential for sharing data across country boundaries for research purposes and for clinical care received abroad

  - Continue to refine and improve common data formats and quality control processes.

- **Refining regulatory frameworks:**
  - Enable patients and providers to trace who uses their data and for what purpose (including automated notifications), and demonstrate commitment to enforcing serious sanctions and penalties for data misuse, with ongoing public debate around this topic where needed to maintain trust
  - Make fast-track pathways in the regulatory approval system available for specific certain cases
  - Continue to publicize data-driven improvements to care quality and efficiency.
- **Building capabilities:**
  - Embed statistics, informatics and data science as part of all education programs, from primary to tertiary education
  - Create training programs and specific innovation grants for research that brings together data science and healthcare, with a focus on collaboration with clinicians
  - Fund programs to train junior academics for data science roles in the healthcare industry.

Health systems delivering quick wins and setting these clear strategic priorities will be better able to exploit the potential of data science and AI. Public engagement and communication must continue throughout the process. The public's valid concerns over the sharing and use of their data must be acknowledged. However, the benefits of data science and AI in health are too great to risk inaction.

GLOSSARY

Algorithm

A set of instructions that are followed to reach an outcome, decision or recommendation.

Artificial intelligence (AI)

A broad term encompassing machine learning, computer vision, natural language processing (NLP), virtual assistants and automated robotics. The science of making computers learn from data and interact with the human world.

Big data

Very large data sets that exceed the storage capacity or processing power of traditional tools. Big data "represents the information assets characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value,"[74] and its prevalence is growing as the ability to collect large amounts of data in real time from users or consumers of products and services increases.

Blockchain

A blockchain is a decentralized computer network designed to perform transactions that are incorruptible, irreversible, transparent and verifiable, and redundantly stored. This technique allows for clear audit trails in data – each interaction with data is appended to the chain, cannot be removed, and in a manner where any tampering with the chain is detectable.

Computer vision

The science of processing, interpreting, labeling and extracting information from images. This might range from analyzing MRI scans to measuring blood-flow rates to processing the input to an autonomous vehicle.

Data science

The methods, processes and systems used to analyze, understand and draw insights from large and complex data sets.

Deep learning

A machine-learning method that consists of multiple layers of nonlinear processing units. Each layer uses the output of the previous layer as an input. This hierarchical structure encourages the formation of multiscale representations of the system, naturally extracting relevant features from the data.

Governance

The institutional configuration of legal, professional and behavioral norms of conduct, conventions and practices that, taken together, govern the collection, storage, use and transfer of data and the institutional mechanisms by and through which those norms are established and enforced.

Machine learning

Techniques to allow a machine to learn approaches to solving problems directly from the input data, without being explicitly programmed or using specific models for the task.

Metadata

'Data about data', contains information about a data set. For example, this information could include why and how the original data was generated, who created it and when. Metadata may also be technical, describing the original data's structure, licensing terms and the standards it conforms to.

Natural language processing (NLP)

Tools that can understand, process, extract or output information as intelligible human language. Examples include Twitter chatbots and Amazon's Alexa home assistant app.

Neural network

A machine-learning method using a series of simple, nonlinear processing units, taking inspiration from biological neurons. Neural networks can describe arbitrary nonlinear functions, but their design makes them easy to fit on large data sets.

Personally identifiable information

Information that could be used to identify an individual, or assign a particular record to a specific person. This includes names, dates of births, health record IDs, social security numbers and images of faces. The use and dissemination of identifiable information is protected to varying degrees by data protection laws. Most tools and devices that use medical information can be developed using anonymized data sets, in which identifiable information has been removed.

Reinforcement learning

A technique for training machines that operate within dynamic environments. Rather than ascribing values or penalties to every action the machine makes (for example, a move in chess), the only feedback given is a measure of cumulative reward (the outcome of the game). As it learns, the machine must balance the need to explore new situations against the benefit of using past experience.

Reinforcement learning agents can learn just by their own process of exploration. Deepmind's AlphaGo Zero recently became the world's best player at the ancient Chinese game of Go, just by playing against itself.[75]

Sensitive data

Data that individuals would not wish to be shared – for example, salary information or disease diagnoses. Precise legal definitions vary by jurisdiction.

Supervised learning

An approach for training machine-learning systems on data sets where the desired output of the system has been labeled (often by a human) in at least some cases. The machine learns to predict the labeled outputs as best it can from the other data available.

Unsupervised learning

The task of learning the structure of a data set that has no categorizations externally imposed on it or clear output variables assigned to it. One purpose is the extraction of natural features to describe or represent the data – for example, identification of clusters of data points.

ACKNOWLEDGMENTS

The Forum advisory board for this report was chaired by Professor Aziz Sheikh, Professor of Primary Care Research & Development and Director, Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh and Co-Director, NHS Digital Academy.

This report was written by Professor Aziz Sheikh in collaboration with:

- Dr Giles Colclough, Data Science Specialist, McKinsey & Company
- Grail Dorling, Health Specialist, McKinsey & Company
- Dr Farhad Riahi, Assistant Professor, McGill University Faculty of Medicine, and former Partner, McKinsey & Company
- Dr Saira Ghafur, Lead for Digital Health Policy, Institute of Global Health Innovation, Imperial College and Honorary Consultant in Respiratory Medicine, Imperial College Healthcare NHS Trust

Sincere thanks are extended to the members of the advisory board who contributed their unique insights to this report:

- Dr David Bates, Chief of General Internal Medicine, Brigham and Women's Hospital
- Dr David Blumenthal, President, The Commonwealth Fund
- Rachel Dunscombe, CEO and Director of Digital, NHS Digital and Northern Care Alliance Group
- Ahmed K Elmagarmid, Executive Director, Qatar Computing Research Institute
- Dr John Halamka, Chief Information Officer, Beth Israel Deaconess Medical Center
- Dr Adam Hill, Chief Medical & Strategy Officer of Oncimmune
- Marten Kaevats, National Digital Advisor, Government of Estonia
- Dr Dominic King, Medical Director, DeepMind Health
- Professor Henrique Martins, Chairman of the Board, Shared Services of the Ministry of Health (SPMS)

- Professor Andrew Morris, CEO, Health Data Research UK (HDRUK)
- Dr Veli Stroetmann, Head of eHealth Research and Policy, empirica Communication and Technology Research
- Professor Mihaela van der Schaar, Man Professor of Quantitative Finance, Oxford University
- Professor Effy Vayena, Health Ethics and Policy Lab, Department of Health Sciences and Technology, ETH Zurich

The interviews that informed this report were conducted by Giles Colclough and Grail Dorling of McKinsey & Company. The chair and authors thank all who contributed. Any errors or omissions remain the responsibility of the authors.

REFERENCES

01. Henke N et al. The age of analytics: Competing in a data-driven world. McKinsey Global Institute, 2016. Available at: www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world [Accessed 30 June 2018].

02. Henke N et al. The age of analytics: Competing in a data-driven world. McKinsey Global Institute; 2016. Available at: www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world [Accessed 30 June 2018].

03. Markoff J. A learning advance in artificial intelligence rivals human abilities. *The New York Times*, 10 December 2015. Available at: www.nytimes.com/2015/12/11/science/an-advance-in-artificial-intelligence-rivals-human-vision-abilities.html [Accessed 30 June 2018].

04. Borowiec S. Google's AlphaGo AI defeats human in first game of Go contest. *The Guardian*, 9 March 2016. Available at: www.theguardian.com/technology/2016/mar/09/google-deepmind-alphago-ai-defeats-human-lee-sedol-first-game-go-contest [Accessed 30 June 2018].

05. Bughin J et al. *Artificial intelligence: The next digital frontier?* McKinsey Global Institute, 2017.

06. Pentland A et al. Big data and health: Revolutionizing medicine and public health. Doha, Qatar: World Innovation Summit for Health (WISH), 2013. Available at: www.wish.org.qa/wp-content/uploads/2018/01/27425_WISH_BigData_Report_web.pdf [Accessed 30 June 2018].

07. IBM Marketing Cloud. 10 key marketing trends for 2017. 2017. Available at: public.dhe.ibm.com/common/ssi/ecm/wr/en/wrl12345usen/watson-customer-engagement-watson-marketing-wr-other-papers-and-reports-wrl12345usen-20170719.pdf [Accessed 30 June 2018].

08. McCallum JC. Memory prices (1957–2017) – data compiled from various sources. Available at: www.jcmit.net/memoryprice.htm [Accessed 30 June 2018].

09. Europa eHealth website. SNS Portal transforms access to public healthcare in Portugal. Available at: joinup.ec.europa.eu/document/sns-portal-transforms-access-public-healthcare-portugal [Accessed 30 June 2018].

10. Bates DW et al. Why policymakers should care about 'Big Data' in healthcare. *Health Policy and Technology*, 2018; 7(2): 211–216. Available at: www.sciencedirect.com/science/article/pii/S2211883718300996 [Accessed 30 June 2018].

11. YouGov. International views on healthcare, online survey. Doha, Qatar: World Innovation Summit for Health (WISH), June 2018. Unpublished.

12. YouGov. International views on healthcare, online survey. Doha, Qatar: World Innovation Summit for Health (WISH), June 2018. Unpublished.

13. YouGov. International views on healthcare, online survey. Doha, Qatar: World Innovation Summit for Health (WISH), June 2018. Unpublished.

14. Rock Health. 50 things we now know about digital health consumers: Results from 2016 digital health consumer adoption survey. 2016. Available at: rockhealth.com/reports/digital-health-consumer-adoption-2016 [Accessed 30 June 2018].

15. Sciencewise. Big data: Public views on the collection, sharing and use of personal data by government and companies. 2014. Available at: webarchive.nationalarchives.gov.uk/20170110135457/http:/www.sciencewise-erc.org.uk/cms/assets/Uploads/SocialIntelligenceBigData.pdf [Accessed 30 June 2018].

16. YouGov. International views on healthcare, online survey. Doha, Qatar: World Innovation Summit for Health (WISH), June 2018. Unpublished.

17. Henke N et al. The age of analytics: Competing in a data-driven world. McKinsey Global Institute, 2016. Available at: www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world [Accessed 30 June 2018].

18. Humber River Hospital Foundation. www.hrhfoundation.ca [Accessed 30 June 2018]. (Unverified commercial claim).

19. Narasimhan V. Three things that will change medicine in 2018. World Economic Forum Annual Meeting, 2018. Available at: www.weforum.org/agenda/2018/01/3-things-change-medicine-2018-big-data-healthcare [Accessed 30 June 2018] (unverified commercial claim).

20. Bughin J et al. Artificial intelligence: The next digital frontier. McKinsey Global Institute Discussion Paper, 2017. Available at: www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx [Accessed 30 June 2018] (unverified commercial claim).

21. G Miller et al. Quantifying national spending on wellness and prevention. *Advances in Health Economics and Health Services Research*, 2008; 19: 1–24.

22. Flores M et al. P4 medicine: How systems medicine will transform the healthcare sector and society. *Personalized Medicine*, 2013; 10(6): 565–576.

23. Bughin J et al. Artificial intelligence: The next digital frontier. McKinsey Global Institute Discussion Paper, 2017. Available at: www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx [Accessed 30 June 2018] (unverified commercial claim).

24. IBM. Watson at work. 2018. Available at: www-05.ibm.com/innovation/uk/watson/watson_in_healthcare.shtml [Accessed 30 June 2018] (unverified expert opinion).

25. Dickson B. Feeling sick? Consult your AI chatbot. *PCMag UK*, 13 September 2017. Available at: uk.pcmag.com/health-fitness/91098/feature/feeling-sick-consult-your-ai-chatbot [Accessed 29 May 2018].

26. Price L. GYANT: AI enabled triage service. Healthcare.digital, 2018. Available at: www.healthcare.digital/single-post/2018/01/09/GYANT-AI-enabled-triage-service [Accessed 30 June 2018].

27. McKinsey & Company. *A future that works: Automation, employment and productivity*, McKinsey Global Institute; 2017.

28. Henke N et al. The age of analytics: Competing in a data-driven world. McKinsey Global Institute, 2016. Available at: www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world [Accessed 30 June 2018].

29. Sheikh A et al. Implementation and adoption of nationwide electronic health records in secondary care in England: Final qualitative results from prospective national evaluation in "early adopter" hospitals. *British Medical Journal*, 2011; 17(343): d6054.

30. Robertson A et al. The rise and fall of England's National Programme for IT. *Journal of the Royal Society of Medicine*, 2011.

31. Tang P et al. Personal health records: Definitions, benefits, and strategies for overcoming barriers to adoption. *Journal of the American Medical Informatics Association*, 2006; 13(2): 121–126.

32. Open Banking Ltd. www.openbanking.org.uk [Accessed 30 June 2018].

33. Kuzniewicz M et al. A quantitative, risk-based approach to the management of neonatal early-onset sepsis. *Journal of the American Medical Association: Pediatrics*, 2017; 171(4): 365–371.

34. Flatiron Health publications. Available at: flatiron.com/publications [Accessed 10 April 2018].

35. Nwaru BI et al. Can learning health systems help organisations deliver personalised care? *BMC Medicine*, 2017; 15: 177.

36. Hawcutt DB et al. Susceptibility to corticosteroid-induced adrenal suppression: a genome-wide association study. *The Lancet Respiratory Medicine*, 2018; 6(6): 442–450.

37. Serviço Nacional de Saúde (SNS). 2018. Available at: www.sns.gov.pt/cidadao/siga-sns-sistema-integrado-de-gestao-do-acesso-no-servico-nacional-de-saude [Accessed 30 June 2018].

38. Serviço Nacional de Saúde (SNS). 2018. Available at: www.sns.gov.pt/apps/te-m-s-tempos-medios-na-saude [Accessed 30 June 2018].

39. Maps available at: tempos.min-saude.pt/#/instituicao/211 [Accessed 30 June 2018].

40. Diário da República Eletrónico (DRE); 2018. Available at: dre.pt/web/guest/pesquisa/-/search/106938486/details/maximized [Accessed 30 June 2018].

41. Artificial Intelligence in Medical Epidemiology (AIME). Dengue outbreak prediction platform. 2018. Available at: aime.life [Accessed 30 June 2018] (unverified commercial claim).

42. McKinsey & Company. Automation case studies: A provocative view of future work settings. McKinsey Global Institute, 2015.

43. University of Pittsburgh Medical Center (UPMC). UPMC announces $2B investment to build 3 digitally based specialty hospitals backed by world-leading innovative, translational science. Available at: www.upmc.com/media/NewsReleases/2017/Pages/upmc-innovates-announcement.aspx [Accessed 30 June 2018].

44. Nwaru BI et al. Can learning health systems help organisations deliver personalised care? *BMC Medicine*, 2017; 15: 177.

45. Friedman CP et al. Toward an information infrastructure for global health improvement. *Yearbook of Medical Informatics*, 2017; 26: 16–23. Available at: www.ncbi.nlm.nih.gov/pmc/articles/PMC5623976/#CR57 [Accessed 12 August 2018].

46. BRAINS (Brain Imaging in Normal Subjects) Expert Working Group. Improving data availability for brain image biobanking in healthy subjects: Practice-based suggestions from an international multidisciplinary working group. *NeuroImage*; 2017; 153: 399–409.

47. Cresswell KM et al. Why every health care organization needs a data science strategy, *New England Journal of Medicine Catalyst*, 22 March 2017. Available at: catalyst.nejm.org/healthcare-needs-data-science-strategy [Accessed 6 July 2018].

48. Smart W. Lessons learned review from the WannaCry ransomeware cyber attack. NHS England, 2018. Available at: www.england.nhs.uk/wp-content/uploads/2018/02/lessons-learned-review-wannacry-ransomware-cyber-attack-cio-review.pdf [Accessed 30 June 2018].

49. Sood HS et al. Has the time come for a unique patient identifier for the US? *New England Journal of Medicine Catalyst*, 21 February 2018. Available at: catalyst.nejm.org/time-unique-patient-identifiers-us [Accessed 30 June 2018].

50. DeepMind Health. Trust, confidence and verifiable data audit. 2018. Available at: deepmind.com/blog/trust-confidence-verifiable-data-audit [Accessed 30 June 2018].

51. Zhang Q et al. Forecasting seasonal influenza fusing digital indicators and a mechanistic disease model. *Proceedings of the 26th International Conference on World Wide Web*, 2017: 311–319.

52. Bousquet J et al. Assessment of thunderstorm-induced asthma using Google Trends, *The Journal of Allergy and Clinical Immunology*, 2017; 140(3): 891–893. Available at: www.jacionline.org/article/S0091-6749(17)30855-2/fulltext [Accessed 30 June 2018]

53. Digital Imaging and Communication in Medicine (DICOM). DICOM standards, 2018. Available at: www.dicomstandard.org/current [Accessed 30 June 2018].

54. OSCAR EMR. Website, 2018. Available at: oscar-emr.com [Accessed 30 June 2018].

55. Cresswell K et al. Establishing data-intensive healthcare: The case of hospital electronic prescribing and medicines administration systems in Scotland. *Journal of Innovation in Health Informatics*, 2016; 23(3): 842.

56. Robertson AR. Tightrope walking towards maximising secondary uses of digitised health data: a qualitative study. *Journal of Innovation in Health Informatics*, 2016; 12,23(3): 847.

57. Sheehan M. Broad consent is informed consent. *British Medical Journal*, 2011; 343: d6900.

58. UK Biobank. UK Biobank participant consent form, version 20061124; 2006.

59. Qatar Biobank. Qatar Biobank participant information booklet; 2013.

60. US Department of Health & Human Services. Attachment C – recommendations for broad consent guidance. 2017. Available at: www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-c-august-2-2017/index.html [Accessed 30 June 2018]

61. US Food and Drug Administration. Expedited Access Pathway Program. 2017. Available at: www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/ucm441467.htm [Accessed 30 June 2018].

62. Heller N. Estonia, the digital republic. *The New Yorker*, 18 December 2017.

63. Fernandez-Luque L et al. Erratum to: Implementing 360° quantified self for childhood obesity: Feasibility study and experiences from a weight loss camp in Qatar. *BMC Medical Informatics and Decision Making*, 2017; 17 (1). Haddadi H et al. 360 quantified self. Queen Mary University of London, 2015. Available at: qmro.qmul.ac.uk/xmlui/bitstream/handle/123456789/13473/Haddadi%20360%20Quantified%20Self%202016%20Accepted.pdf?sequence=1 [Accessed 30 June 2018].

64. Sathyanarayana A et al. Impact of physical activity on sleep: A deep learning based exploration. 2016. Available at: www.researchgate.net/publication/305638109_Impact_of_Physical_Activity_on_SleepA_Deep_Learning_Based_Exploration.

65. Interview with Dr John Halamka, CIO, Beth Israel Deaconess HealthCare, 23 February 2018.

66. Interview with Dr John Halamka, CIO, Beth Israel Deaconess HealthCare, 23 February 2018.

67. Wang D. Deep learning for identifying metastatic breast cancer. ArXiv, 2016. 1606.05718v1.

68. Lung Nodule Analysis (LUNA). Results of the LUNA 2016 grand challenge. Available at: luna16.grand-challenge.org/results [Accessed 30 June 2018].

69. Esteva A et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017; 542: 115–118.

70. Xiao E. Alibaba Cloud doubles down on healthcare for its AI business, *TechInAsia*, 30 March 2017. Available at: www.techinasia.com/alibaba-cloud-et-medical-brain [Accessed 30 June 2018] (unverified commercial claim).

71. Verily Life Sciences. Eyes: A window into heart health. 19 February 2018. Available at: blog.verily.com/2018/02/eyes-window-into-heart-health.html [Accessed 30 June 2018].

72. The Farr Institute of Health Informatics Research. 100 ways of using data to make lives better. 2018. Available at: www.farrinstitute.org/public-engagement-involvement/100-ways-of-using-data-to-make-lives-better [Accessed 30 June 2018].

73. University of Edinburgh. Data-driven innovation. 2018. Available at: www.ed.ac.uk/local/city-region-deal/data-driven-innovation [Accessed 30 June 2018].

74. De Mauro A et al. A formal definition of big data based on its essential features. *Library Review*, 2016; 65: 122–135.

75. Silver D et al. Mastering the game of Go without human knowledge. *Nature*, 2017; 550: 354–359.

WISH RESEARCH PARTNERS

State of Qatar
MINISTRY OF PUBLIC HEALTH

WISH gratefully acknowledges the support of the Ministry of Public Health

THE AGA KHAN UNIVERSITY

المستشفى الأهلي
AL-AHLI HOSPITAL

Apollo HOSPITALS

Ballarat Health Services
Putting your health first®

THE BEHAVIOURAL INSIGHTS TEAM

Burnet Institute
Medical Research. Practical Action.

CDDEP CENTER FOR Disease Dynamics, Economics & Policy
WASHINGTON DC • NEW DELHI

Centre intégré universitaire de santé et de services sociaux de l'Estrie – Centre hospitalier universitaire de Sherbrooke
Québec

CHATHAM HOUSE
The Royal Institute of International Affairs

Deloitte Centre *for* Health Solutions

engagedIN

جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY
عضو في مؤسسة قطر
Member of Qatar Foundation

مؤسسة حمد الطبية
Hamad Medical Corporation
HEALTH • EDUCATION • RESEARCH صحة • تعليم • بحوث

醫院管理局
HOSPITAL AUTHORITY

Imperial College London
Institute of Global Health Innovation

NHS
Imperial College Healthcare
NHS Trust

Implemental
improving mental health and wellbeing around the world

KPMG

mind
for better mental health

Moorfields Eye Hospital NHS
NHS Foundation Trust

NHS
National Institute for Health Research

Nursing now

QATAR FUND FOR DEVELOPMENT
صندوق قطر للتنمية

الهلال الأحمر القطري
Qatar Red Crescent

QCRI
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

مركز دراسات التشريع الإسلامي والأخلاق
Research Center for Islamic Legislation and Ethics
Member of Hamad Bin Khalifa University

سدرة للطب
Sidra Medicine

ST VINCENT'S HEALTH AUSTRALIA

جامعة كالجاري في قطر
UNIVERSITY OF CALGARY IN QATAR

UCL

THE UNIVERSITY of EDINBURGH

NHS
University Hospitals of Leicester
NHS Trust

Vancouver CoastalHealth
Promoting wellness. Ensuring care.

Waitemata
District Health Board
Best Care for Everyone

ISBN 978-1-912865-01-7

www.wish.org.qa