Realising the value of health care data: a framework for the future

EY
Building a better working world

How can we place a value on health care data?

A framework proposal focused on the UK's health care ecosystem and patient records held across the National Health Service (NHS)

Medicine is no longer a clinical science just supported by data, it's moving to a field defined as clinical science in collaboration with data science. Patient data is one of the most important drivers of this change. Unlocking the insights contained in patient genomic and phenotypic data is of high value to all the key stakeholders in the health care ecosystem: patients, providers, payers and the life sciences sector. In this paper, we show the methodological tools that can be used to estimate the value of patient data, specifically the data held by the UK's NHS. We show how realising the value of this data can help the UK Government achieve its health priorities, and place the UK at the forefront of global health care innovation.

Contents

Executive summary

Unlocking the power of health care data to fuel innovation in medical research and improve patient care is at the heart of today's health care revolution. When curated or consolidated into a single longitudinal data set, patient-level records will trace a complete story of a patient's health, wellness, diagnosis, treatments, medical procedures and outcomes. Health care providers need to recognise patient data for what it is: a valuable intangible asset desired by multiple stakeholders, a treasure trove of information.

Among the universe of providers holding significant data assets, the UK's NHS is the single largest integrated health care provider in the world, its patient records covering the entire UK population from birth through to death. There are two primary approaches to quantifying the value of data:

1. A market-based approach, calculating the implied "per record" valuation multiples of comparable data assets or valuation multiples of companies with significant patient data assets

2. An income-based approach, which quantifies value based on the economic benefit to be generated from the curated data set

Applying these approaches, we estimate that the 55 million patient records held by the NHS today may have an indicative market value of several billion pounds to a commercial organisation. We estimate also that the value of the curated NHS data set could be as much as £5bn per annum and deliver around £4.6bn of benefit to patients per annum – generated through potential operational savings for the NHS, enhanced patient outcomes and creation of wider economic benefits to the UK, generated through 'big data', artificial intelligence and personalised medicine.

The curated NHS data set is an intangible asset with a current valuation of several billion pounds and a realisation of £9.6bn per annum in benefits that could be unlocked following the generation of insights.

There will be a significant process and technology costs associated with aggregation, cleaning, curating, hosting, analysing and protecting the transformation of these raw data records into a consolidated longitudinal patient-level data set. The costs associated with this data transformation requires further research and clearly could impact data set valuation. To ensure success the NHS and the UK government will need to partner with companies that can help unlock these valuable patient insights.

It is critical that the analyses and innovations adhere with medical ethics and research regulations. Patients have to be informed and need to be confident that their data is being used for their own and public good, and that their privacy and rights are safeguarded.

Ultimately, analysis and insights generated from this unique NHS data set can help the UK Government achieve its health priorities on prevention, care and costs, place the NHS and the UK at the forefront of health care innovation, and make the NHS the envy of the world.

1 Patient data as a desirable intangible asset

Unlocking the power of health care data to fuel innovation in medical research and improve patient care is at the heart of today's health care revolution. It is made possible by advances in health care technologies and data digitalisation, and achieved through the analysis of real-world evidence contained within a given patient's medical records. When curated or consolidated into a single longitudinal data set, these patient-level records will trace a complete story of a patient's health, wellness, diagnosis, treatments, medical procedures and outcomes.

Unlocking insights from such medical records would be of high value and benefit to all key stakeholders in the health care ecosystem — patients, health care providers, payers, pharmaceutical companies and medical device manufacturers. Consider a few examples:

- Deeper disease understanding
- Treatment effectiveness and safety or pharmacovigilance
- Increased in the quality of care, such as faster and early diagnosis
- Observation of real-world patient outcomes and clinical pathway efficiency
- Improved patient access to therapies
- Evidence of cost-effectiveness and outcomes to inform value-based payments
- Efficient target identification for new treatments and medicines
- Clinical trial design for target populations and reduced time to market for new therapies
- Operational and cost-effectiveness of health care delivery, and workforce planning
- Enablement of personalised medicine — right treatments for the right patients
- Socioeconomic impacts of health, wellness and health care

Given these considerations, health care providers need to recognise the patient data they accumulate for what it truly is — a valuable intangible asset desired by multiple stakeholders, and a treasure trove of information detailing health, wellness, illness and the associated care pathways.

Realising the value of NHS data

Amongst the providers holding significant data assets, the UK's NHS is the single largest integrated health care provider in the world. Its patient records covering the entire UK population from birth to death is an estimated 55 million lives.

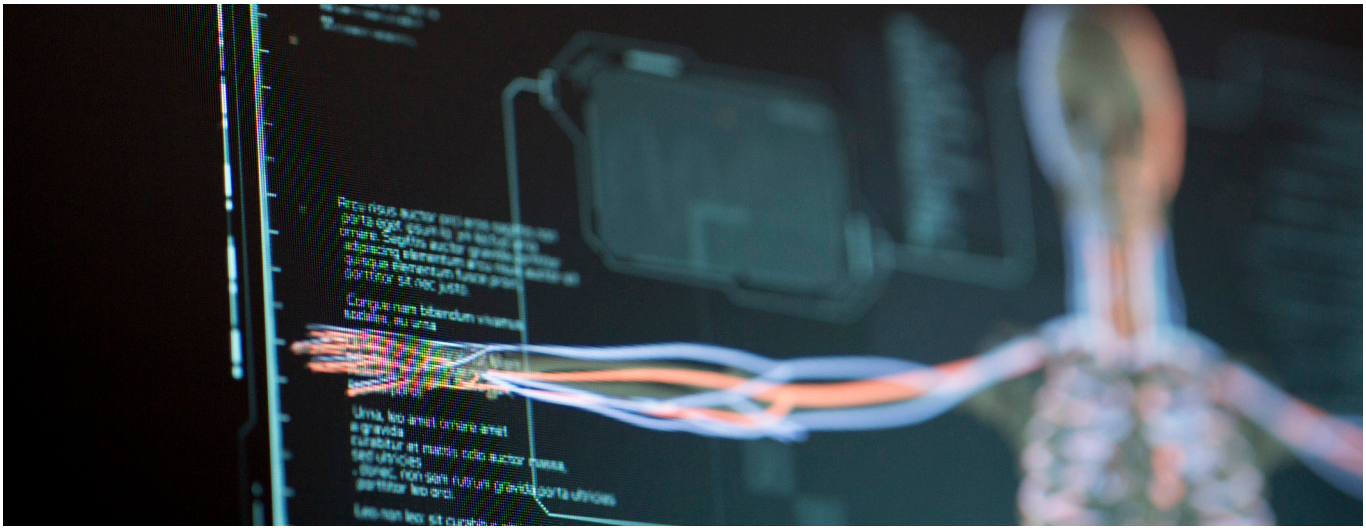
Ultimately, the analytical insights derived from this unique data set can help the UK Government achieve its health priorities on prevention, care and costs. In the near term, monetising access to the data set can yield a compelling revenue stream for the NHS — especially until other countries amass and put on the market their digitised patient records of comparable depth and scale.

We estimate that the 55 million patient records[*] held by the NHS today may have an indicative market value of several billion pounds to a commercial organisation.

This figure is based on benchmarks derived from the following:

- The trading values of health and life sciences companies with significant patient data assets
- Recent M&A transactions involving such companies and data

[*] A patient record is defined as quantitative or qualitative information about a patient of a single patient's medical history and care held by the NHS in any format. The term 'patient record' includes electronic patient records, electronic medical records, records of episodic care and supplementary registries such as genomic or other 'omic' profiling with or without associated medical histories.



This estimate will likely increase as the proportion of genomic records in the NHS data set increases, with sequencing costs falling over time. Increases will occur as the volume records available to curate and link grows. We estimate that in the next 5 years the total number of patient records associated whole genome sequences will increase from the current 100,000 to over 5 million. In addition to this increase in genomic profiles, the UK's Department of Health and Social Care has announced that patients will be asked to give consent for their genome data to be securely analysed by approved researchers.

We expect additional value to be unlocked from patient records by emerging scientific and medical innovations. These include panomic** profiling where advances in molecular biological technologies are leading to detailed mapping of a patient's panomic profile.

To unlock further value of its patient records, the NHS would need to create longitudinal patient-level data sets by combining all care settings (e.g., primary, secondary and social care) and available genomic profile data for each patient. It should be noted, however, that there would be a significant cost associated with this data transformation.

We estimate that the value of the curated NHS data set could be as much as £5bn per annum, delivering around £4.6bn worth of benefit to patients per annum, if this effort is undertaken. The value to patients would come from potential operational savings for the NHS, enhanced patient outcomes and the generation of wider economic benefits to the UK.

The worth of data — three questions to consider

Such considerations raise the pressing tactical question of how one places a value on patient health data — NHS-owned or any other. Quantifying this value is a complex endeavour dependent on multiple variables:

- Each major stakeholder from across the health care ecosystem will have a different view of what value is — its perceived value would not necessarily be commercial in nature.
- There are significant costs that need to be taken into account associated with curating, processing and analysing electronic patient records.
- Challenges around interoperability and diverse data contents need to be addressed.

As a starting point, one needs to explore the following three stage-setting questions:

1. What is each stakeholder group's key objective or the issue it needs to resolve?
2. How does each stakeholder intend to use the data set?
3. Is the proposed use ethical and in line with the Government's regulatory guidance?

Once these perspective-setting questions are considered, two approaches can be used to quantify the data's value. Both approaches focus on the fundamental drivers of health care data value.

** Panomics refers to a range of molecular biology technologies including:
1. Proteomics — the analysis and quantification of proteins
2. Metabolomics — the quantification of multiple small molecule types, such as amino acids
3. Microbiomics — the examination of all the microorganisms of a community simultaneously
4. Epigenomics — the characterisation of reversible modifications of DNA
5. Transcriptomics — the examination of RNA levels genome-wide
6. WGS — Whole Genome Sequencing

2 Drivers of health care data value

The health care industry and health care providers have historically generated vast amounts of data as a result of clinical and regulatory requirements. However, the collection and storage of this data was generally in paper form. Recently, the accessibility and amount of digitalised data has exploded, growing at an exponential rate.

Though all data can have value, this will vary depending on its characteristics. To measure the monetary value of data, one must first understand how these different characteristics increase or decrease its value.

On the basis of our experience and the review of published literature on the subject, the characteristics that affect the value of health care data can be broadly categorised into four main pillars as presented in figure 1. Please refer to the appendices for more details.

Advanced data mining, such as artificial intelligence (AI) and natural language processing, will help reveal patterns and trends that will allow health care providers and other stakeholders to improve diagnoses, treatments and operational effectiveness.

In fact, the understanding of the associations and trends within health data is already being used to study health at both the population level, and the personalised, individual level. Companies are utilising health and data analytics to build predictive models that identify and predict high-risk patients, and are developing early intervention plans to improve health and care proactively.

Analytics and predictive models are helping address one of the biggest costs to health care – the treatment of chronic diseases. The main factors that influence the risk of developing chronic diseases include medical history, demographic or socioeconomic profiles and comorbidities. Thus, by aggregating data related to these factors, predictive analytics can help identify those individuals who are at risk of chronic conditions.

> “Having genetic information and longitudinal data allows us to paint the clearest picture on patient epidemiology, progression, and overall experience.
>
> **Director, Field Health Outcomes, Pharmaceutical company**

Figure 1. Framework for analysing characteristics that impact the value of a data set

1	Nature	▸ Data type (patient, payer, product, provider and scientific research) ▸ Data availability or time frame (contemporaneous vs. historical with time lag) ▸ Exclusivity or scarcity (available from a single source vs. multiple) ▸ Granularity or detail (aggregated vs. transaction level) ▸ Source or seller (original source/generator of the data vs. reseller)
2	Data quality, maturity and embedded analytic insight	▸ Raw (unorganised with potential data gaps and inconsistencies) ▸ Curated (i.e., organised and easy to work with) ▸ Aggregated longitudinally for the same patient or record ▸ Analysed with descriptive statistics, insights and predictions or forecasts provided
3	Complexity of data capture	▸ Source or the party generating the data ▸ Accessibility of data (open source vs. paid) ▸ Data capture (auto-captured vs. collected with human intervention)
4	Use/application	▸ Use and potential impact ▸ Exclusivity (exclusive licence vs. data being offered to multiple buyers) ▸ Limitations on use ▸ Usage by other businesses or competitors

3 Valuation approaches to consider

The value of patient data is proportional to its content, and comes from the value of insights and innovations that can be unlocked from the curated data.

We see two primary approaches to quantifying this value.

1

A 'top-down' or market-based approach

This approach estimates the value of a data set based on its profile – size, content, nature, availability, maturity, uniqueness and the quality of the data – and on the observed market pricing of functionally comparable data assets. Market pricing is analysed by calculating the implied 'per-record' valuation multiples of comparable data assets or, in the absence of such information for 'pure' data assets, valuation multiples of companies with significant patient data assets. By applying the benchmark per-record valuation multiples to the NHS data set, one can estimate the data's value based on recent 'market transactions'.

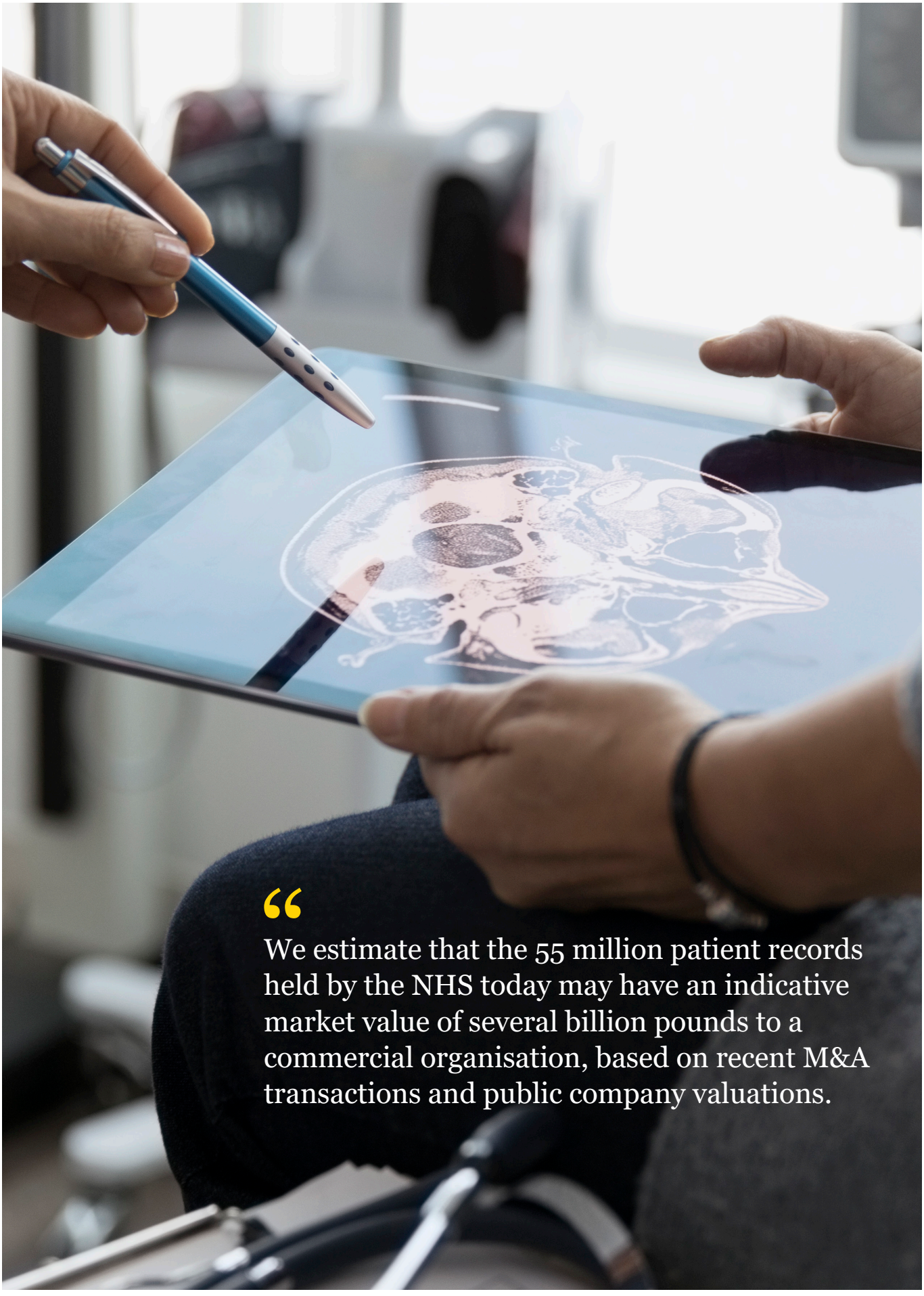
2

A 'bottom-up' or income-based approach

This approach quantifies value based on the economic benefit to be generated from the curated data set under its proposed use. Value can be expressed in terms of economic uplift and commercial opportunity:

Economic uplift: The benefits to patients include faster and more accurate diagnosis, improved outcomes and new products, improved care pathways, and operational efficiency. Benefits to the UK economy include socioeconomic benefits or increased gross value added.

Commercial opportunity: This refers to the incremental profit or licensing income generated from a successful business deployment of insights gained from the data. For example, if the insights are deployed by a life sciences company, such business deployment may translate into reduced drug development time and the launch of a novel therapy.



> We estimate that the 55 million patient records held by the NHS today may have an indicative market value of several billion pounds to a commercial organisation, based on recent M&A transactions and public company valuations.

4 Application of the market-based approach

Simply put, the application of the market-based approach calls for a thoughtful analysis of two matters:

1. Strengths and weaknesses of the subject data set relative to comparable data sets that are being traded in the market
2. Per-record valuation multiples inferred from comparable market transactions or company valuations

The application of this framework is illustrated in figure 2, with a focus on the NHS data set.

Analysing the NHS data set's strengths and weaknesses — the power of longitudinal data

Each fragmented and isolated data set will have a certain value, albeit low. True value is generated when these data sets are aggregated, curated, processed and linked to create a longitudinal data set.

The transformed data set will contain a patient's unique phenotypical and genomic information that has been captured over a period of time. Panomic data generated from patient's tissue and fluid samples will become increasingly available over the next few years as molecular biology technologies advance. Greater value can be unlocked when the data set undergoes analysis using basic or advanced analytical methodologies to identify commercially valuable insights. A range of diagnostic, descriptive, predictive and prescriptive analytical methodologies should be applied to maximise insight generation and extract as much value as possible. It is important to bear in mind that there will be a significant process and technology costs associated with aggregation, cleaning, curating, hosting, analysing and protecting these transformed data sets.

To quantify the rise in value, one should then compare the characteristics of the longitudinal data set with those that have been the target of recent transactions or public company trading activity within the relevant data subcategory — e.g., primary care, secondary care or genomic profile. This can be achieved by reviewing publicly available data, private databases, company reports and press releases.

Understanding the NHS data and its ecosystem

We assume that there are 55 million longitudinal primary care records or electronic health records (EHRs) in the UK. These are held by the patients' general practitioner (GP) surgery where the patients are registered and seen regularly.

In the ideal world, every patient would have a single EHR, all records would be digital and centrally housed, and all health care platforms and data sets would be standardised and interoperable. But in reality, the NHS health care data ecosystem is made up of multiple fragmented and geographically dispersed data silos. Each data set contains multiple data types and formats.

These isolated data sets are often organised at the patient, population and organisation levels rather than solely at the patient level. The data contents and complexity vary, as does the data quality. Many records contain both structured and unstructured data detailing demographics, diagnosis, procedures, medical images, treatments, outcomes and genomic data.

The extent to which these patient records are fully digitalised has to be established, as many of the earlier patient records will be non-digital and maintained within paper files, with attached test results, medical images and x-rays, or stapled electrocardiogram (ECG) cardiac traces.

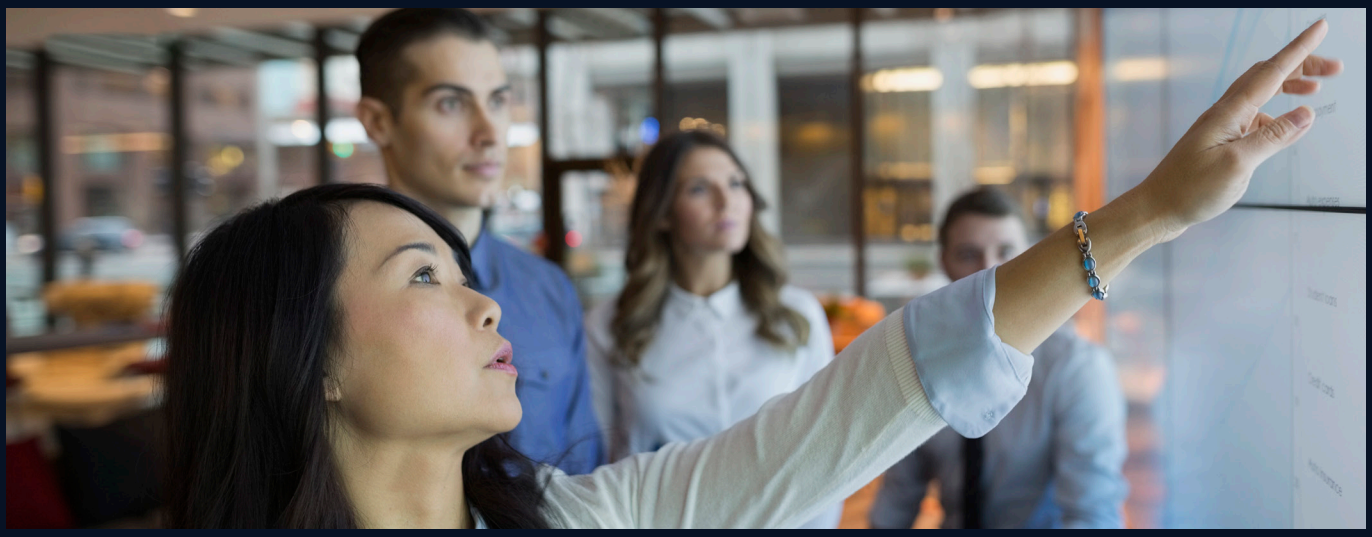
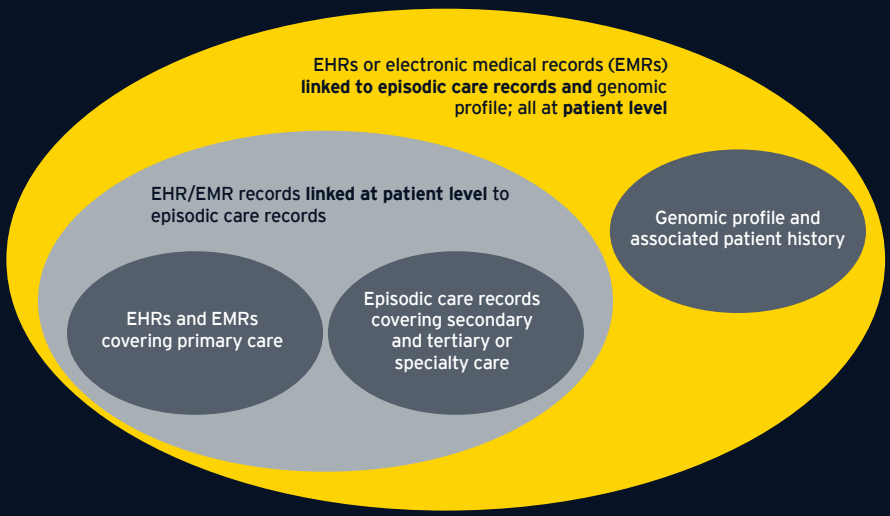


Figure 2. Framework for estimating the value of the NHS data set using the market-based approach

1. Understand the NHS data set
2. Characterise and evaluate individual and linked data set using qualitative framework with the following categories
3. Analyse comparable market transaction and select valuation multiples
4. Apply valuation multiples to NHS data set
5. Estimate NHS data set value and identify necessary limitations



Central NHS data sets (local, regional and national level) include prescriptions, medicine expenditure, supply chain, procurement, financial, resourcing and workflow or process details, and for local authorities, data sets will include the details of social care

We must assume that in 2019, most patients will have an EHR, although how far back the data is digitalised is a key question. Higher-level and less detailed extracts are held by the local clinical commissioning groups (CCGs) and these are used primarily for health and financial reporting.

In addition to the 55 million primary care records, there are approximately 23 million patient-episodic care records or electronic medical records (EMRs); these are captured when patients receive secondary or tertiary (or specialist) care, and are held by the hospital trusts. These EMRs will include details on admissions and medical interventions in accident and emergency departments.

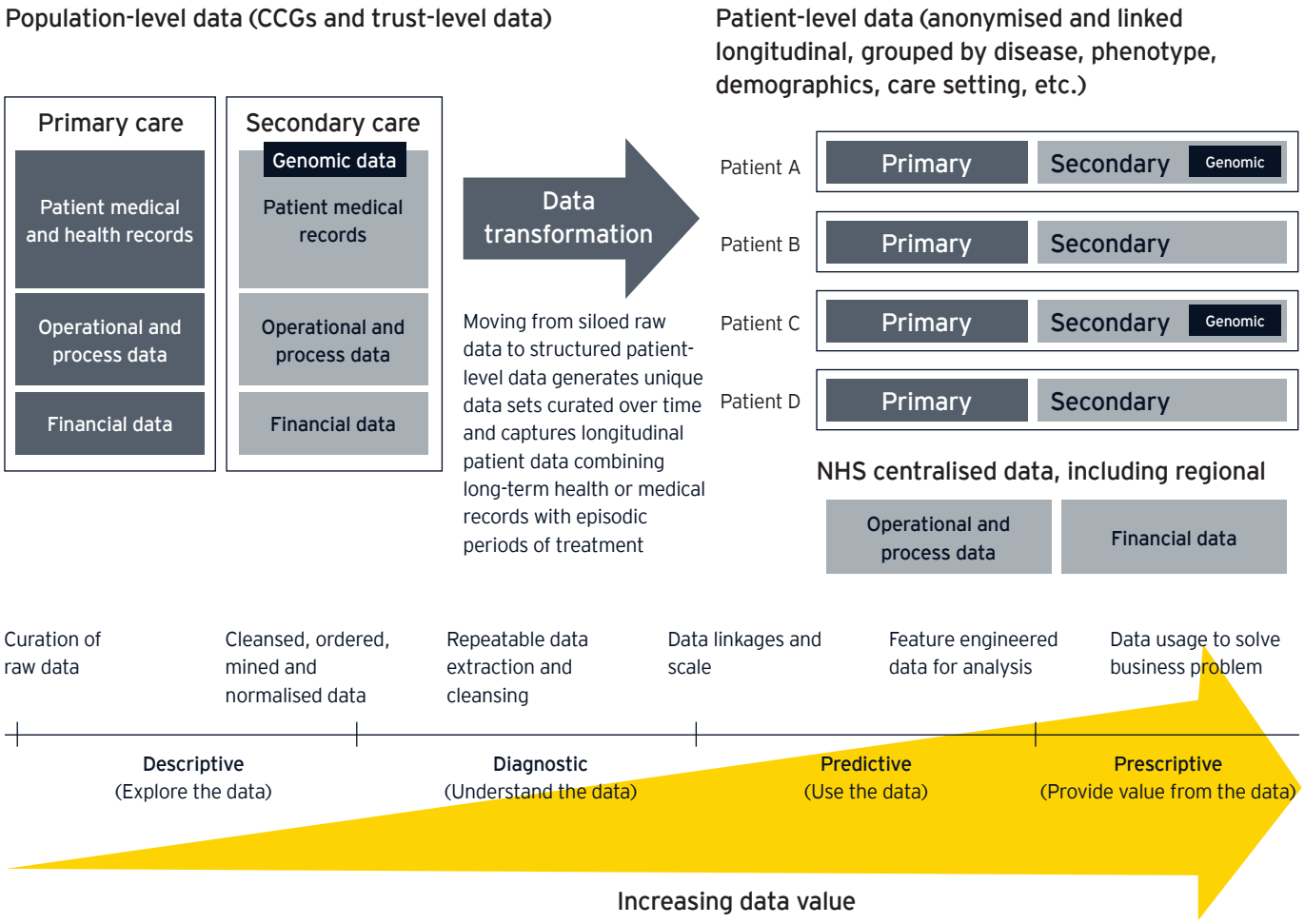
There are separate records that detail social care that patients are receiving and are collated by local authorities from the social care providers.

Data sets (both digital and paper) are held by NHS data controllers – e.g., NHS Trusts for secondary or tertiary care and by CCGs or GPs for primary care.

Other health care-related data sets are created for specific needs – e.g., Genomics England has sequenced the DNA of 100,000 patients with cancer, rare diseases and infectious diseases, and collate these sequences for the purpose of understanding the disease, earlier

Figure 3. Transformation of population-level data to patient-level data

The health ecosystem is made up of multiple types of real-world data (patient-, population- and organisation-level data). Moving from siloed raw data to structured patient-level data generates unique data sets curated over time and captures longitudinal patient data combining long-term health or medical records with episodic periods of treatment. These can be analyses to generate insights and provide value to patients, providers, payers, pharmaceutical companies and the wider economy as a whole.



> “There is little commercial use of genomic data on its own, but it becomes powerful when wedded to clinical outcomes. For example, the combination of EMR and genomic data is essential for biomarker identification, enhancing the predictive power of models.
>
> **Director, Roche, Sequencing Solutions**

diagnosis and, ultimately, personalised care. In addition to data controllers, there are also data processors – e.g., commissioning support units (CSUs), the Office for National Statistics and Public Health England. Their role is to provide population-level analyses of health, wellness, disease and cost or operation effectiveness. There are some examples where the data processors may be both controllers and processors – e.g., NHS Digital, CCGs, local authorities and social care providers.

Curating and linking fragmented patient data sets into a single longitudinal patient record

For each patient, there is both health and medical information contained in EHRs (detailing primary care) and EMRs (detailing episodes of secondary or tertiary care). This is illustrated in figure 3. Recently, genomic profiling (genotype or whole genome sequencing) is becoming more frequent as sequencing costs are reduced. This has the potential to vastly increase complexity of analyses and the data set size. Additional information from registries and social care may also be available. Linking data sets, both geographically and longitudinally, along the patient pathway offers the potential of a significant increase in the number and value of insights that can be derived to improve patient care and accelerate medical research.

Interoperability between health care solutions and data platforms is pivotal to success. Part of the challenge (and cost) of unlocking value is developing the most suitable processes and data workflows to aggregate, clean and convert these fragmented and isolated data records into a single high-quality, analysable data set. Protection of these data files is vital and an appropriate cybersecurity solution is critical.

One should also be aware of the increase in complexity (and cost) of the longitudinal data when records and episodes of care are linked; and that the operational risks of updating and refreshing the vast data set increases accordingly.

In the case of NHS records, resolving the operational challenge of linking the raw data is aided by the availability of unique patient identifier numbers. The entire population of the UK has a unique 10-digit patient identifier, the NHS number, that stays with individuals from birth until death. This NHS number will be associated with any care the patient receives in their lifetime, and as such will cover primary, secondary, and tertiary health and medical records; any social care received; and any data collected through health-related registries that patients have volunteered information to – e.g., by participating in Genomics England's 100,000 Genomes project.

Per-record valuation multiples – analysing comparable market transactions or company valuations

A number of public and private companies with assets that include patient data can be identified by a detailed review of publicly available and transactional databases, and from these the implied or observed pricing for patient data can be analysed. The implied business valuation multiples on a per-record basis can be calculated for data they own or can access. These valuations can be aligned to the data categories we described and illustrated in the framework above in figure 2.

Caution must be taken in analysing transactions – the EHR or EMR data is often acquired as part of a broader set of capabilities, such as advanced analytics, including AI. Hence, isolating the value of data becomes a challenge and there will be differences in the nature, quality, complexity and application of data in each data set and with each transaction.

Electronic patient records (EHR and EMR)
From our review, it is clear that the EHR or EMR data, in isolation, has lower value than genomic data. Recent transactional activity in these patient data categories has been motivated by acquirers seeking to broaden their capabilities. The range of estimated values per patient record and company-level analyses are listed in the appendix – supplement.

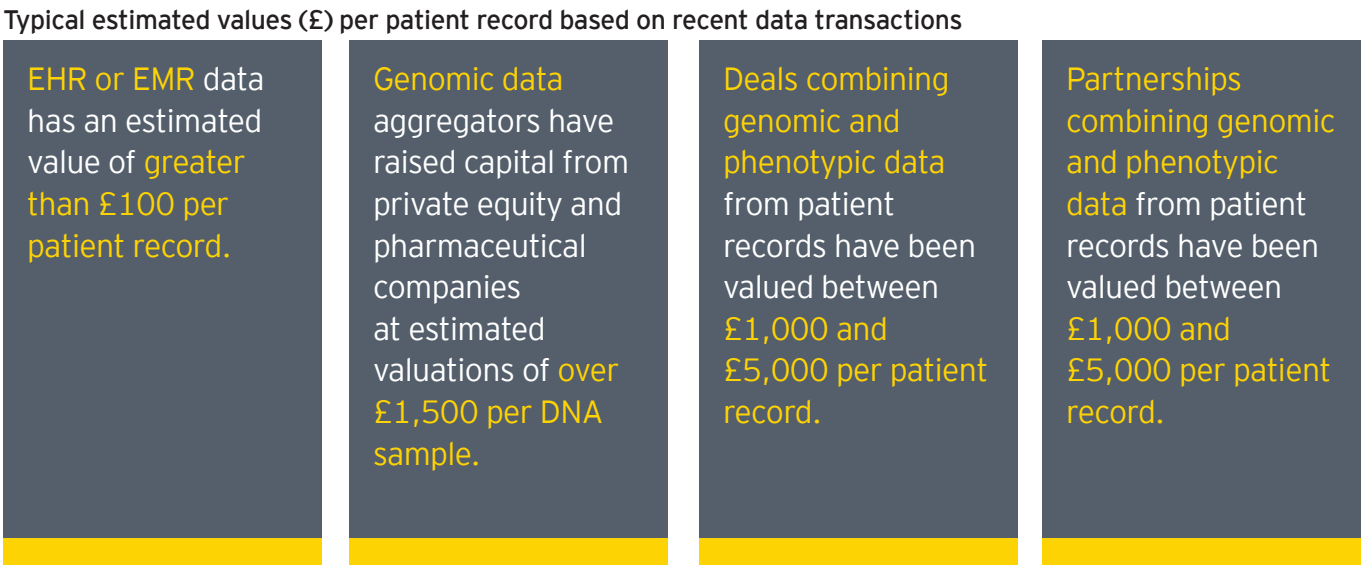
Examples where we are unable to quantify the implied data value vs. other capabilities, include the IMS Health-Quintiles merger in 2016, which included 15 petabytes of prescription, promotional, claims and approximately 530 million patients' records or EHR data, leading to the combined entity (IQVIA) having a market capital of approximately US$20bn and 50,000 employees.

Again, Roche's acquisition of Flatiron was a combination of patient data, an EHR and an oncology platform that brings all available data together for review by the oncology board. In this case, Roche acquired real-world data or evidence, collated from Flatiron's network of approximately 280 oncology communities, which provided patients' records of approximately two million oncology practices. Based purely on the acquisition cost and number of records, one can estimate a value of US$950 per patient record; however, the data content is of the premium quality, and the value includes access to Flatiron's health care technology, which connects oncologists on a shared platform.

In some cases, the investment information is not in the public domain, such as the Great Point Partners' acquisition of Corrona, which held an observational registry with data on approximately 65,000 patients.

We can make estimates from other deals where EHR data sets were acquired for testing and refining AI solutions. IBM spent approximately US$4bn on acquisitions of Phytel, Explorys, Merge and Truven in order to enhance its Watson Health (AI) offerings. The individual acquisitions allow us to estimate that IBM paid Merge, e.g., approximately US$30 per patient record for an anonymised pool of approximately 30 billion medical images, including x-rays, computerised axial tomography (CAT) and magnetic resonance imaging (MRI) scans.

Figure 4. Observed values (£) per patient record based on recent data transactions (January 2019)



Genomic profiles and medical histories

Genomic data is recognised by experts as a key component for future drug discovery and also personalised medicine. However, this data is of limited value in isolation without longitudinal health and medical data.

There is currently a scarcity of large data sets due to the high historical costs of sequencing and this scarcity is driving the valuation of per deoxyribonucleic acid (DNA) sample to be less than US$1500 (i.e., each individual patient profile). Some examples are national-level initiatives, such as the UK Biobank and Genomics England. These are often open-access for academic researchers and clinicians to analyse as well as contribute to, but require a subscription fee when pharmaceutical companies request access. An example of this is the nine months' exclusive data access granted by UK Biobank to Regeneron and GSK: the first data set has just been released to health researchers from this collaboration offering an unprecedented 'big data' resource to enhance understanding of human biology and aid in therapeutic discovery. Details of the estimated record value are described in the paragraph below. The range of estimated values per patient record and company-level analyses are listed in the appendix — supplement.

Observed valuations of consumer-facing genomic profiling providers, such as 23andMe and Ancestry.com provide useful valuation benchmarks for genomic data. It should be noted that these companies mainly analyse genome samples rather than whole genome sequences and the genetic analysis offered combines genetic profile and the persons' associated health and family history (note, this does not include EHR or EMR data).

With this caveat, Silverlake's and GIC's investment in Ancestry.com indicates an estimated value of approximately US$1700 per genotype record, based on 1.5 million DNA samples sourced from consumers and linked to more than 20 billion digitised historical records and over 80 million family trees. The research partnership between 23andMe and Genentech indicates an approximate value of US$5,000 per patient record, highlighting that a premium is placed on more 'medicalised' consumer genotype profiles with associated medical history (in this case, focusing on Parkinson's disease). The 23andMe data was sourced directly from over one million consumers, with 23andMe recruiting and profiling the genotypes of 12,000 Parkinson's patients (and 1,300 parents and siblings), while a subset of 3,000 Parkinson's patients samples underwent whole genome profiling.

Panomic profile and medical histories

Panomics, which analyses patient tissue and body samples in more detailed data than genomics, is gaining traction across the pharmaceutical industry to expedite drug discovery and focus on personalised medicines. Evotec and Celgene have entered into a third long-term strategic drug discovery and development partnership in the field of targeted protein degradation. In this collaboration, Evotec and Celgene will leverage Evotec's panomics platform in order to identify drug targets. There is currently no available data on panomics record transactions and our value estimation for the NHS data set does not include value that could be unlocked by curation of patient records with panomic profiles.

Consolidated genomic profile and electronic patient records (EHR and EMR)

There are relatively few historical examples of genotypic and phenotypic data being linked, and this provides the NHS with a unique opportunity. The NHS holds EHR and EMR records from all care settings and has the ability to link these to the Genomics England 100,000 Genomes Project.

Transactions or partnerships for these linked data sets command high valuations per patient record (approximately US$1,000–US$5,000 per record) as they provide an holistic perspective on patient cohorts. These longitudinal phenotypic and genotypic patient-level data sets are inherently valuable for drug discovery and wider population health analysis due to the scale and depth of data provided for analysis. An example of this is Roche's acquisitions in 2018, which gave it analytic capabilities and ownership of phenotypic and genotype data for its oncology franchise.

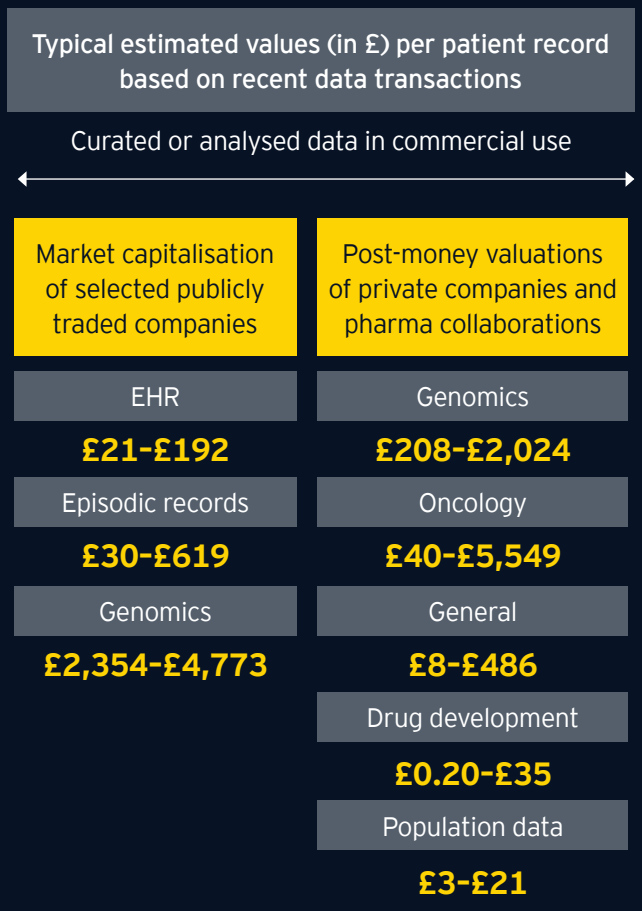
Examples where we can provide an estimated value for any transaction or partnership investment, include Amgen's success in the public bidding held by the Icelandic Parliament to create a genealogical database linking genotypic to phenotypic data. Data was sourced from the Icelandic population and contained 600 genetic samples, from a population of 320,000, and we can estimate a value of approximately US$1,300 per record for this partnership.

A second example is Genomics Medicine Ireland (GMI) which has a project objective to undertake one of the world's largest whole genome sequencing programme, creating 600 new jobs in five years; here, we estimate a value of US$1,000 per record. The partnership between Regeneron and GSK to sequence 500,000 participants in the UK biobank project (with nine months of exclusive access to the sequenced data) yields an estimate of US$3,000 per record.

One final example is the investments that lead to Roche's acquisition of Flatiron as it demonstrated how value increases as phenotypic and genotype data sets are combined, with a further value-add when additional

Figure 5. Summary of the estimate ranges of market value for NHS data sets

Benchmark valuation multiples are estimated primarily using the observed valuations of companies with assets that include patient data. The data is understood to be curated (clean) and actively used. The range of estimated values per patient record and company-level analyses are listed in the appendix – supplement. (Analyses as of January 2019).



services are added on top (such as health care platforms to collect patient data and advanced analytics capabilities). Foundation Medicine and Flatiron jointly launched a clinico-genomic database following the integration of Foundation Medicine's genomic profiling assays with Flatiron's EHR outcomes data, building an initial database of approximately 20,000 patients as a subset of the 400,000 cancer patients in the EHR database. The objective was to help researchers and biopharmaceutical partners accelerate drug development in oncology. Foundation Medicine and Flatiron were both acquired by Roche at a cost which an estimate of US$6,000 per record for Foundation Medicine genomic data and US$950 per record for Flatiron EHR data. This is based on the price Roche paid for data, oncology EHRs, health care platform and analytical services, which was approximately US$2.4bn for Foundation Medicine and US$2.6bn for Flatiron, which would equate to US$250, per clinico-genomic record.

Considerations to take into account whilst estimating the value of patient records held by the NHS

For the UK's NHS data set, we have made the assumption that the patient records fall into the following categories and in each, we know the approximate number:

- Around 55 million longitudinal primary care records
- Around 23 million episodic specialty care patient records
- Around 100,000 DNA codes of patients with cancer, rare diseases and infectious diseases collated by Genomics England

For a simple estimate, we would take the number of patient records in each category and multiply by the relevant estimated value for each patient record (on the basis of recent data transactions described in figure 5 and the text above). However, we need to take into account that the value of these data sets increases when they are linked. There should be :

- Approximately 23 million linked records where the patient's primary and specialty care records are curated
- Approximately 100,000 linked patient records that detail primary care, episodic specialty care and genomic profiles based on overlaps and our ability to link using the patient's NHS number (for this, a premium value per record should be applied)

As mentioned above, we have listed down two factors to consider for a simple estimate by stating 'should' as completed digital records for some patients may be unavailable, or there could be missing data. It is essential that when estimating the total value of the patient records held by the NHS, one should consider the following factors

affecting the value of patient data (these are described in more detail in the appendix – supplement):

1. Available digitised patient records – not all patient historical data will be digitalised, and an estimate of the period that corresponds to the EHR or EMR must be made
2. Nature of the data, including demographics, health, wellness, diagnosis, treatments, medical procedures and outcomes
3. Data access availability, availability of substitutes or alternative data sources, granularity and levels of detail
4. Data quality, maturity and embedded analytical insights (factors contributing to this include whether the data is raw, unstructured or structured, curated, quality assured, aggregated across individual records, aggregated longitudinally at the patient-level, or analysed by descriptive statistics or predictive analytics)
5. Overlap across three data categories and synergy value
   - Amongst the 55 million patient records, most patients would have received multiple episodes of secondary or tertiary care.
   - But 23 million patients or fewer would have patient-level data sets with both primary and secondary or tertiary care since there are only 23 million episodic care records.
   - Lastly, fewer than 100,000 records will contain genomic and phenotypic data, because there currently only 100,000 patients who has undergone genome sequencing; this number is expected to rise to five million in five years.
6. Variability in benchmark valuation multiples (analysis in figure 5 above was done by relying on business valuations as opposed to raw data asset valuations; i.e., they rely on valuations of public and private companies that possess significant patient data assets)
7. Understanding the minimal statistically significant data sample size needed for research and unlocking value from the patient records
8. Diminishing returns on increasing data set size (additional improvements in data algorithms and corresponding extraction of insights will start becoming marginal in nature once the user begins exceeding the minimal statistically significant data sample size milestone)

Additionally, a key consideration when estimating value is the uniqueness of the data: are similar data sets available from other sources, or countries? The more unique the data, the greater will be the value, especially if a commercial option is also desirable.

Finally, caution must be taken when directly applying fair market values for each data category. As noted previously, often when patient data is acquired, especially for data contained within EHR or EMR, the transactional deal will often include acquisition of additional capabilities such advanced analytical or health care platforms in addition to the curated patient data.

5 Application of the income approach

The NHS, through its analysis of the patient-level data it holds, can unlock significant operational savings, enhanced patient outcomes, and wider economic benefits through the application of 'big data', AI and personalised medicine. If the NHS were to curate these data sets and generate longitudinal patient-level records, it would be able to unlock significant value, although there would be a significant cost associated with this data transformation and generation of insights.

We estimate that the value to the NHS could be as much as £5bn per annum and deliver around £4.6bn per annum of benefit to patients. The value to the NHS can come from operational savings for the NHS, enhanced patient outcomes and generation of a wider economic benefits to the UK. The economic analyses reveals a number of short-, medium- and long-term impacts which yield financial and economic benefits as a result of more cost-effective and informed spending.

Framework for valuing patient data based on economic benefits to patient, the NHS and the UK economy

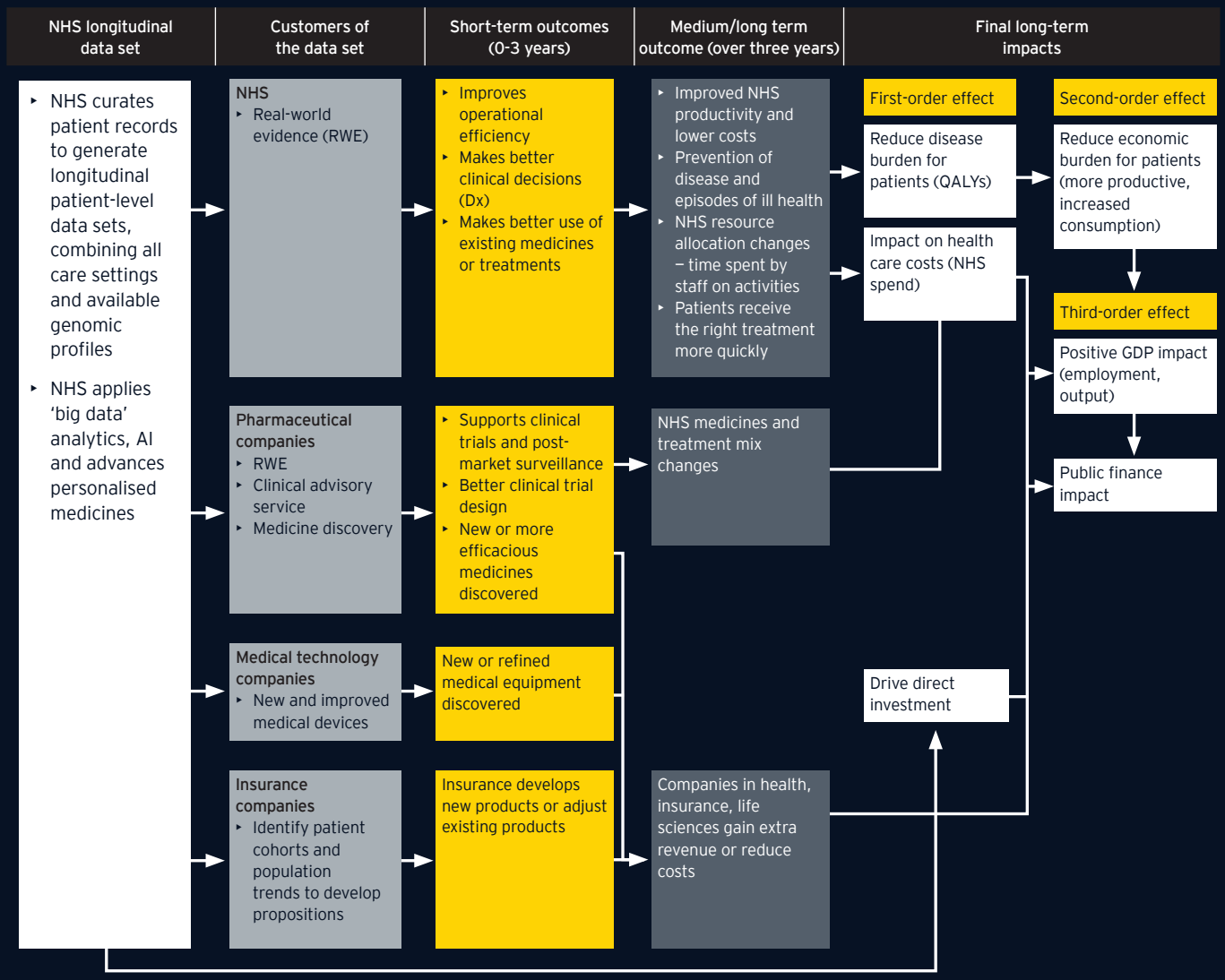
Estimating the economic value (i.e., the benefits) that would be generated from the curation, processing and analyses of a large data set provides an alternative approach to assessing the value of data. It also allows us to include the wider benefits to the economy in addition to patient benefits and economic uplift for the NHS. This approach generates both qualitative and quantitative outputs. It also takes into account the costs of transformation of the data sets, as the services and technology costs will all contribute to the gross value added (GVA) for the local economy. The longitudinal patient-level data set and full use of use of 'big data' will enable better decision-making and an ability to explore AI and further advancements in personalised medicine.

The economic uplift approach is an evidence-based review methodology that draws on a combination of expert consultation, literature review and assessment of working papers which enable the development of an impact framework to underpin economic modelling and analysis (figure 6). These analyses are split into three categories of impact — 'big data', AI and personalised medicine — and takes into account the 'customers' of the insights generated: in this case, the NHS and three industry sectors (i.e., pharmaceutical, medical technology and insurance companies). The outputs are estimated for each of the 'beneficiaries' of the insights: patients, the NHS and the wider economy. It is important to consider the realisation of the economic benefits as each of the three categories has differential impact across short-term (less than three years) and medium- or long-term (greater than three years) timeframes for each of these categories (see figure 7).

'Big data' are large, static data sets that can be analysed to provide real-world insights on trends and patterns. Uniting all NHS data sets into a single analysable patient- level data set would create opportunities to improve NHS productivity and delivery of care. Our analyses estimate that productivity savings could be equivalent to 2% of the annual NHS budget

Figure 6. Schematic of impact framework for estimating the economic benefits and uplift

This impact framework demonstrates a number of short-, medium- and long-term impacts which yield financial and economic benefits as a result of more cost-effective and informed spending (QALY – quality-adjusted life year; GDP – gross domestic product).



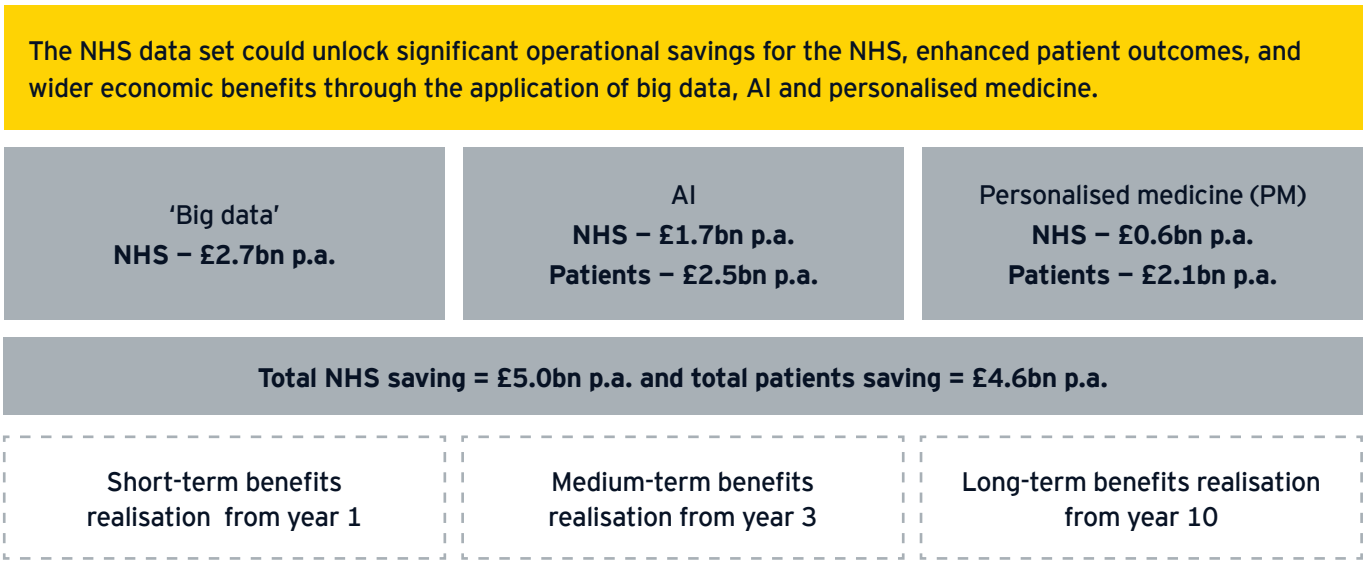
and equate to £2.7bn per annum in productivity savings. Moreover, these savings could start to be realised as soon as one year after NHS data set curation (figure 6). The NHS could use a big NHS data set to identify best practice care pathways to be adopted. This would improve NHS productivity and make its limited resources go further. A unified NHS data set would be necessary to achieve integrated care, currently a policy goal, according to several influential voices in the health sector. The existence of this unique data set could attract investment from life sciences sector in terms of research and development activities. Patients and the wider economy are also likely to benefit from better health and better care through improved pathways, and more R&D spending in the UK, but there is, to date, limited evidence on the scale of these potential benefits.

The full potential of a single longitudinal patient-level NHS data set can be realised following development and curation with the application of **AI**. AI here means the application of computational processes that mimic human intelligence (e.g., reasoning, learning and adaptation) to big, real-time data sets. AI has been used in health care for early detection, diagnosis, decision-making, treatment and research. AI could support the NHS to work more quickly and accurately. Our economic modelling reveals that the economic benefits to the NHS of applying AI to the longitudinal patient-level NHS data set could be realised in the medium term (from three years onwards) with savings to the NHS of £1.7bn per annum and benefits to patients of £2.5bn per annum (figure 6).

These benefits will continue long into the future as the data set grows, and the breadth and depth of analyses continue to reveal insights and suggest solutions to help the NHS address clinical and financial problems. Early adoptions of AI within the health care ecosystems have been promising. Applying AI to the NHS data set will yield insights that can be used to improve patient outcomes through improved monitoring and better use of medicines. As a follow-on from this, AI can help reduce errors, provide better diagnostics and improve demand planning. Ultimately, improved patient health will also have implications for the UK economy and consequent positive impacts on public finances. AI could be applied to the NHS data set to make its operations more efficient and more accurate, including reducing adverse reactions to medicines, improving the speed and accuracy of diagnostics, and, from an operational perspective, leading to more effective and accurate theatre scheduling and capacity planning.

Figure 7. Summary of economic value to the NHS benefit to patients

The value would be in terms of operational savings for the NHS, enhanced patient outcomes and generation of a wider economic benefits to the UK.



Personalised medicine tailors treatments to the characteristics of a particular patient. Personalised medicine will be enabled if the NHS creates a patient-level longitudinal data set containing both phenotypic and genotype-linked data. Genomic data can be used for drug discovery, to give more precise diagnoses, make more efficient use of medicines, and increase the quality and length of life. Through its promise of more targeted and effective treatment, personalised medicine could have significant impacts on the NHS, patients, the life sciences industry and the economy. Our analyses estimate that the economic benefits of personalised medicine could be £0.6bn per annum in savings for the NHS, with a patient benefit of £2.1bn per annum (figure 6). The benefits of personalised medicine would be realised only in the longer term, over a period of 10 years or more.

The NHS would be able to dedicate fewer resources to ineffective treatments for non-responders, and patients should be treated more efficiently with medicines or technologies identified as likely to be effective for them. Early evidence indicates that personalised medicine could both save the NHS from wasting money on ineffective treatments, and lead to improved morbidity and reduced mortality. Life sciences companies could benefit through revenue gains from introducing new medicines. There may also be benefits from clinical trials that are becoming faster and more cost-effective by better targeting of participants through genomics. By helping pharmaceutical companies to target the right participants for clinical trials, the NHS could reduce the length and cost of trials in the UK. Evidence indicates that better targeting could reduce the time it takes to get a medicine to market by 1.7 years. Moreover, a healthier and more productive workforce would have economy-wide benefits and positive public finance implications. The GVA benefits to the UK from a healthier and larger workforce could be as high as £1bn per annum.

Realisation of the economic benefits. The application of 'big data', AI and PM will require a level of set up and initial investment for the benefits to realise. The economic analyses highlight that 'big data' will materialise first, followed by AI, and then PM last but with potentially the most significant long-term impact (figure 7).

Figure 8. Realisation of the economic benefits for 'big data', AI and personalised medicine resulting from the creation of a NHS longitudinal patient-level data set and the analytical identification of insights

	Big data	AI	PM
Short term (less than 3 years)	NHS bodies procure technology they need	Data set created and leads to useful insights on best practice	Pharma begins process of developing new medicines
Medium term (3-9 years)	Best practices and learnings spread through NHS	Best practices in short run implemented across NHS	New medicines go through clinical trials
Long term (greater than 10 years)	All potential benefits realised	Continued discovery of useful insights	New medicines come to market

6 Concluding comments

The NHS data set holds immense value to all stakeholders across the health care ecosystem — patients, the NHS, the life sciences sector, the UK and society as a whole.

The curated NHS data set is an intangible asset with a current valuation of several billion pounds and a realisation of £9.6bn per annum in benefits (i.e., the NHS benefits worth £5bn per annum and the patient benefits worth £4.6bn per annum) that could be unlocked following the generation of insights.

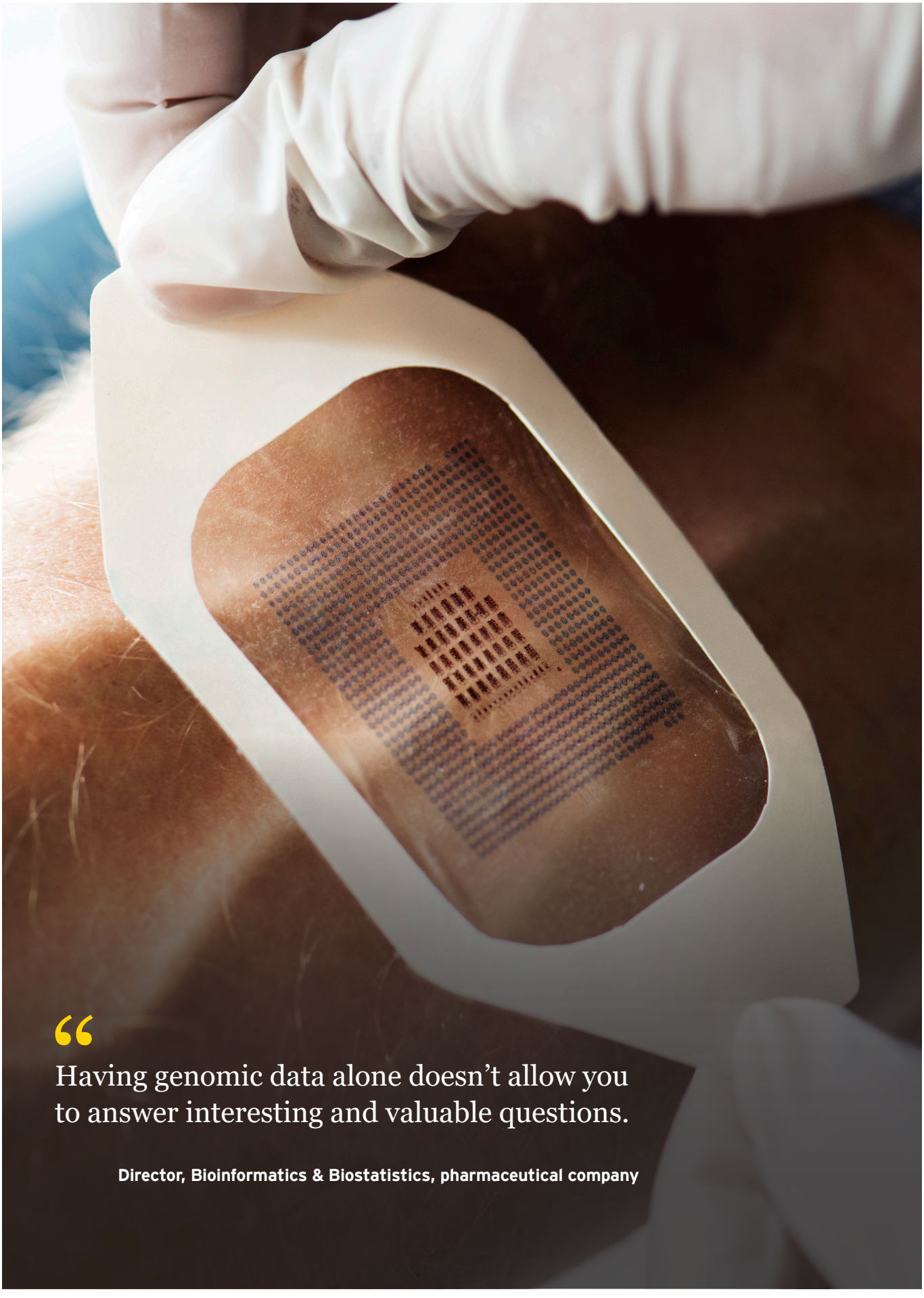
There will be significant process and technology costs associated with aggregating, cleaning, curating, hosting, analysing and protecting the transformation of these raw data records into a consolidated longitudinal patient-level data set. Estimating these costs should be a next step in light of the multifaceted requirements needed to convert isolated NHS patient records into a longitudinal curated patient-level data set. The need and associated costs to address digitalisation and data transformation, cybersecurity, infrastructure as well as solving the interoperability issues must not be under estimated.

A significant and appropriately sized team of highly skilled data scientists, who have access to cutting edge IT and analytical technologies, will be required to generate the insights needed to produce the value as described. The up-front and maintenance costs will come in the form of people, process and technology, and one should not underestimate the scale of the task or its requirements. The NHS and the UK's Department of Health and Social Care will need to partner to ensure success, to access the expertise in technology and clinical analytics needed to unlock valuable insights and drive this transformation.

Analysis of the consolidated NHS data set has the potential to unlock ground breaking data-driven innovations, allow the development of apps for patients and doctors, provide clinical decision support tools supported by intelligent algorithms and advanced analytical methodologies. These innovations have the potential to improve diagnosis, treatment, experience of care, efficiency of the system and overall outcomes for patient, the NHS, public health and the wider health and care system.

It is critical that the analyses and innovations adhere with medical ethics and research regulations. Patients have to be informed and need to be confident that their data is being used for their own and public good, and that their privacy and rights are safeguarded. It is the duty of the NHS to consolidated patient records into the highest quality longitudinal patient level data set, and then to use these data-driven technologies, in a safe, ethical, evidenced and transparent way. This will ensure that the insights and real-world evidence generated can be used confidently to improve the quality and safety of care. Analysis and insights generated from this unique NHS data set can help the UK Government achieve its health priorities on prevention, care and costs.

This health care data transformation can place the NHS and the UK at the forefront of health care innovation.



> Having genomic data alone doesn't allow you to answer interesting and valuable questions.

Director, Bioinformatics & Biostatistics, pharmaceutical company

7 Appendix

Factors affecting value of data

Nature of data

Data access and availability

Data access and availability refer to how quickly data is available to the buyer after it is created. Contemporaneous data is one that is immediately available for analysis.

The age of the data can increase or decrease its value, but in general, rapid access to fresh data is thought to be of more value, as it provides real-time insights and enables timely informed decisions.

Exclusivity or scarcity

When data is only available from a single source, the party holding the data acquires monopoly pricing power over the data. The larger the business disruption caused by losing access to such data, the more significant will be the seller's ability to charge a premium price.

Access to data can be limited via physical barriers, expense of the collection process, business strategy or contractual restrictions.

Granularity of data

In general, the increased granularity or specificity of a data set increases its value by allowing the user to extract additional insights absent from aggregated data sets.

In health care context, granularity translates into patient- and transaction-level data (e.g., medical product dispense, patient encounters, change in status, etc.).

By analysing dispense data at a transaction level, pharmaceutical companies can hone in on dispensing patterns in specific zip codes — which would not be possible using aggregated data.

Another example of value evident in granularity is genome sequencing data. By using predictive analytics on patient-level genome sequencing data, one can identify individual patients that are at high risk for certain conditions. Providers can then take preventative actions with the patient to decrease the likelihood of developing or worsening a medical condition. According to the 2016 McKinsey report: *The Age of Analytics — Competing in a Data-Driven World*, the impact from personalised medicine resulting from such data analytics could result in US$2tn to US$10tn in health care savings on a global basis.

Source or seller

Data that is purchased directly from the party that generated it will generally be more valuable as compared with the data purchased from a reseller. This is due to data accuracy.

Altered or inaccurate data may result in misleading conclusions and faulty decisions. It is presumed that data purchased from a reseller is at a higher risk of having been tampered with or altered, when compared with data purchased from the original source.

It should be noted, however, that many third-party data aggregators and resellers may clean and organise data to make it easier to analyse. The incremental increase of data value resulting from this analytical process is addressed in 5.1.2 of the framework, and should be weighted against the increased risk of data alteration.

Data quality, maturity and embedded analytics insights

Data maturity

The concept of data maturity addresses the progression of a data set along the data analytics process.

More often than not, raw data sets are disorganised and complex, or lack structure. To extract insights from it, one must clean, organise and analyse it.

Complexity of data capture

Data capture

The value of data is also affected by the effort and risk associated with capturing it. In general, data that is auto captured as a by-product of everyday business processes would generally be perceived as less valuable — on a cost-approach basis — than the data that is collected in a separate process with human intervention and by including or not including other incremental resources involved in the process.

Accessibility of data

Generally speaking, any restrictions on access to data that is perceived by multiple parties as useful to their decision-making processes will increase its value. Hence, the value of data that is freely available to any interested party would be lower than the value of the data that calls for certain qualifications from the potential buyer, ranging from a monetary payment or through qualifications, such as citizenship or a need to submit a use case request.

In the case of government or public data sets, a monetary payment set for accessing may not reflect full market value of the data across multiple user groups. When using such payments as price benchmarks in an fair market value (FMV) assessment, careful consideration should be given to understanding the price formation of such access restrictions and their consistency with true market value of the data.

For example, government data, such as census results, allows for ready resource allocation, capital investment planning, policy-making and monitoring numerous other benefits. In the G20 countries, open-sourced data is estimated to be valued between US$700bn and US$950bn per annum according to a 2014 study by Lateral Economics.

Use or application

Use and potential impact

The ultimate determinant of data value is its use and application — its power to inform decisions. These may range from ongoing daily operating decisions of a pharmaceutical company, to reducing time to market for new medicines and improving patient adherence to therapy.

The more uses the data set has to multiple buyers, and the more risky and impactful are the resulting decisions on the user's long-term strategy and operations, the more valuable it is to the buyers.

Factors affecting value of data

Observed or implied value in £ per data record	Q1	Median	Q3	n=
Multiples implied by market capitalisation of selected benchmarks of publicly traded companies				
EHR	21	42	4,990	9
Episodic care records	30	54	39,379	7
Genomics	2,354	2,719	6,826	3
Multiples implied by post-money valuations of private companies and by pharma collaborations				
Genomics	208	1,153	2,338	6
Oncology	40	379	20,905	6
General	8	27	10,255	21
Drug development	0.2	17	51,577	6
Population data	3	9	45	4
Subset of above multiples for companies with combined data sets	4,318	6,297	10,255	2

Appendix – supplement

Company-level indicative estimates of market value – observed or implied price (in £) per patient record

Public companies and EHR (also known as primary care)

Indicative estimates of market value for NHS data sets
Market-based approach: guideline public companies benchmarking
Analysis as of 01 January 2019

#	Company	Ticker	EV	# records	EV/record	Domain
GPCs w/assets that include access to EHR data			[1]	[2]		
1	Computer Programs and Systems, Inc.	NasdaqGS:CPSI	£373,769,930	18,000,000	£21	EHR
2	OptumInsight (UnitedHealth Subsidiary)	NYSE:UNH	£9,118,903,409	216,000,000	£42	EHR
3	Cerner Corporation	NasdaqGS:CERN	£13,336,402,955	100,000,000	£133	EHR
4	AllScripts Healthcare Solutions, Inc.	NasdaqGS:MDRX	£3,074,770,356	16,000,000	£192	EHR
5	athenahealth, Inc.	NasdaqGS:ATHN	£4,249,839,591	106,000,000	£40	EHR
6	NextGen Healthcare, Inc.	NasdaqGS:NXGN	£781,472,941	240,000,000	£3	EHR
7	Winning Health Technology Group	SZSE:300253	£2,321,668,843	n/a	n/a	EHR
8	CompuGroup Medical Societas Europaea	DB:COP	£2,085,836,362	n/a	n/a	EHR
9	B-SOFT Co.,Ltd.	SZSE:300451	£1,076,947,286	n/a	n/a	EHR
10	EMIS Group plc	AIM:EMIS	£548,698,594	40,000,000	£14	EHR
11	Medasys S.A.	ENXTPA:MED	£30,774,801	n/a	n/a	EHR
12	Pharmagest Interactive SA	ENXTPA:PHA	£673,709,001	135,000	£4,990	EHR
13	Alibaba Health Information Technology Limited	SEHK:241	£7,335,594,444	28,000,000	£262	EHR
14	Rayseach Laboratories	OM:RAY B	£300,776,801	n/a	n/a	EHR
	First quartile		£579,951,196	18,000,000	£21	
	Median		£1,581,391,824	40,000,000	£42	
	Third quartile		£3,956,072,282	106,000,000	£192	

Notes:
EV=Estimated value
[1] Source: SP Capital IQ. Amount shown for Optum represents a portion of UnitedHealthcare's EV attributed to Optum Insight by prorata revenue contribution.
[2] Source: Latest 10K filings or equivalent annual report, and/or company websites. For more detail on data source please refer to supporting schedules.

Appendix – supplement (continued)

Public companies and episodic records or EMR (also known as secondary, tertiary or specialty care)

Indicative estimates of market value for NHS data sets

Market-based approach: guideline public companies benchmarking
Analysis as of 01 January 2019

	Company	Ticker	EV	# records	EV/record	Domain
GPCs w/assets that include access to episodes of care or transaction data			[1]	[2]		
1	IQVIA Holdings Inc.	NYSE:IQV	£26,030,825,543	530,000,000	£49	Episodic
2	Inovalon Holdings Inc.	NasdaqGS:INOV	£2,378,296,346	240,000,000	£10	Episodic
3	Medidata Solutions Inc.	NasdaqGS:MDSO	£3,067,675,815	3,800,000	£807	Episodic
4	Tabula Rasa HealthCare Inc.	NasdaqGM:TRHC	£1,040,371,508	n/a	n/a	Episodic
5	Veeva Systems Inc.	NYSE:VEEV	£9,408,263,746	n/a	n/a	Episodic
6	Guardant Health Inc.	NasdaqGS:GH	£2,756,527,429	70,000	£39,379	Episodic
7	Precipio Inc.	NasdaqCM:PRPO	£6,437,121	n/a	n/a	Episodic
8	Syneos Health Inc.	NasdaqGS:SYNH	£5,368,017,071	100,000,000	£54	Episodic
9	WuXi AppTec Co. Ltd.	SHSE:603259	£8,988,144,252	n/a	n/a	Episodic
10	Evolent Health Inc.	NYSE:EVH	£1,162,121,376	2,700,000	£430	Episodic
11	Inovalon Holdings Inc.	NasdaqGS:INOV	£2,378,296,346	240,000,000	£10	Episodic
12	Craneware plc	AIM:CRW	£614,529,127	n/a	n/a	Episodic
	First quartile		£1,131,683,909	3,250,000	£30	
	Median		£2,567,411,888	100,000,000	£54	
	Third quartile		£6,273,048,866	240,000,000	£619	

Notes:
EV=Estimated value
[1] Source: SP Capital IQ. Amount shown for Optum represents a portion of UnitedHealthcare's EV attributed to Optum Insight by prorata revenue contribution.
[2] Source: Latest 10K filings or equivalent annual report, and/or company websites. For more detail on data source please refer to supporting schedules.

Public companies and genomic profiles

Indicative estimates of market value for NHS data sets

Market-based approach: guideline public companies benchmarking
Analysis as of 01 January 2019

	Company	Ticker	EV	# records	EV/record	Domain
GPCs w/assets that include access to genomic data			[1]	[2]		
1	Invitae Corporation	NYSE:NVTA	£598,237,180	220,000	£2,719	Genomics
2	Myriad Genetics Inc.	NasdaqGS:MYGN	£1,785,846,015	n/a	n/a	Genomics
3	Natera Inc.	NasdaqGS:NTRA	£647,694,600	n/a	n/a	Genomics
4	Genomic Health Inc.	NasdaqGS:GHDX	£1,690,872,942	850,000	£1,989	Genomics
5	Exact Sciences Corporation	NasdaqCM:EXAS	£5,720,119,415	n/a	n/a	Genomics
6	Berry Genomics Co. Ltd	SZSE:000710	£1,254,923,166	n/a	n/a	Genomics
7	Illumina Inc.	NasdaqGS:ILMN	£33,902,607,451	n/a	n/a	Genomics
8	OPKO Health Inc.	NasdaqGS:OPK	£1,459,479,999	n/a	n/a	Genomics
9	Twist Bioscience Corporation	NasdaqGS:TWST	£682,625,186	100,000	£6,826	Genomics
	First quartile		£682,625,186	160,000	£2,354	
	Median		£1,459,479,999	220,000	£2,719	
	Third quartile		£1,785,846,015	535,000	£4,773	

Notes:
EV=Estimated value
[1] Source: SP Capital IQ. Amount shown for Optum represents a portion of UnitedHealthcare's EV attributed to Optum Insight by prorata revenue contribution.
[2] Source: Latest 10K filings or equivalent annual report, and/or company websites. For more detail on data source please refer to supporting schedules.

Appendix – supplement (continued)

Private companies w/assets that include access to genomic data

Indicative estimates of market value for NHS data sets
Market-based approach: private companies benchmarking
Analysis as of 01 January 2019

	Target	Buyer	Type of deal	Year	Domain	Deal value	Approx % transacted	Post-money valuation or acquired price	Number of records	Implied EV or record
Private companies w/assets that include access to genomic data								[1]	[2]	
1	23 and Me	GSK	Collaboration	2018	Genomics	£236,658,000	12%	£1,972,150,000	5,000,000	£394
2	NextCODE Health	WuXi PharmaTech	Acquisition	2015	Genomics	£51,275,900	100%	£51,275,900	350,000	£147
3	Helix	DFJ Growth Mayo Clinic	Investment	2018	Genomics	£19,721,500	8%	£257,215,692	n/a	n/a
4	Fabric Genomics	Artis Ventures	Investment	2016	Genomics	£18,143,780	19%	£95,530,946	50,000	£1,911
5	Color Genomics	General Catalyst	Investment	2017	Genomics	£70,555,638	29%	£239,868,660	N/A	N/A
6	DeCODE Genetics	Amgen	Acquisition	2012	Genomics	£327,376,900	100%	£327,376,900	140,000	£2,338
7	DeepGenomics	Khosla Ventures	Investment	2017	Genomics	£10,255,180	78%	£13,173,962	n/a	n/a
8	Genomics plc	Invesco Perpetual	Investment	2018	Genomics	£33,896,328	30%	£113,449,983	800,000	£142
9	Nebula Genomics Inc	ARCH Venture Partners LP	Investment	2018	Genomics	£3,392,098	16%	£20,612,912	10,000	£2,061
10	Verge Genomics Inc.	SUNU Ventures BV	Investment	2018	Genomics	£25,243,520	23%	£108,878,457	n/a	n/a
11	Pathway Genomics Corp.	International Business Machines Corp.	Investment	2015	Genomics	£31,554,400	29%	£107,545,284	n/a	n/a
	First quartile							£73,403,423	72,500	£208
	Median							£108,878,457	245,000	£1,153
	Third quartile							£248,542,176	687,500	£2,024

Notes:
EV=Estimated value
[1] Source: SP Capital IQ, DowJones VentureSource.
[2] Source: EY research and/or company websites.

Appendix – supplement (continued)

Private companies w/assets that include access to oncology data

Indicative estimates of market value for NHS data sets
Market-based approach: private companies benchmarking
Analysis as of 01 January 2019

	Target	Buyer	Type of deal	Year	Domain	Deal value	Approx % transacted	Post-money valuation or acquired price	Number of records	Implied EV or record
Private companies w/assets that include access to oncology data								[1]	[2]	
1	Flatiron	Roche	Acquisition	2018	Oncology	£1,498,834,000	100%	£1,498,834,000	2,200,000	£681
2	Cota Health	IQVIA	Investment	2018	Oncology	£31,554,400	18%	£172,831,337	n/a	n/a
3	Tempus	Baillie Gifford T. Rowe Price Revolution Growth New Enterprise Associates (NEA)	Investment	2018	Oncology	£87,563,460	6%	£1,577,720,000	220,000	£7,171
4	Precision Health AI	SymphonyAI Group	Investment	2017	Oncology	£15,777,200	21%	£75,841,000		£76
5	Paige.AI	Breyer Capital	Investment	2018	Oncology	£19,721,500	21%	£93,251,141	25,000,000	£4
6	Freenome	Andreessen Horowitz	Investment	2017	Oncology	£56,797,920	32%	£175,868,448	n/a	n/a
7	Foundation Medicine	Merck	Acquisition	2015	Oncology	£1,893,264,000	45%	£4,180,958,000	200,000	£20,905
8	Cambridge Cancer Genomics Inc.	AME Cloud Ventures	Investment	2018	Oncology	£3,549,870	15%	£23,665,800	n/a	n/a
9	Mendel Health Inc.	Boostrap Labs	Investment	2017	Oncology	£1,577,720	11%	£13,954,933	n/a	n/a
10	Oncora Medical Inc.	Undisclosed investor	Investment	2017	Oncology	£1,490,945	9%	£16,700,166	n/a	n/a
11	PathAI Inc	8VC	Investment	2017	Oncology	£11,785,568	17%	£70,445,198	n/a	n/a
12	SkinVision BV	PHS Capital B.V.	Investment	2018	Oncology	£5,995,336	19%	£31,775,281	1,156,359	£27
	First quartile							£29,747,911	415,000	£40
	Median							£84,546,071	1,078,180	£379
	Third quartile							£506,609,836	1,939,090	£5,549

Notes:
EV=Estimated value
[1] Source: SP Capital IQ, DowJones VentureSource.
[2] Source: EY research and/or company websites.

Appendix – supplement (continued)

Private companies w/assets that include access to drug development data

Indicative estimates of market value for NHS data sets

Market-based approach: private companies benchmarking
Analysis as of 01 January 2019

	Target	Buyer	Type of deal	Year	Domain	Deal value	Approx % trans acted	Post-money valuation or acquired price	Number of records	Implied EV or record
Private companies w/assets that include access to drug development data								[1]	[2]	
1	twoXAR	SoftBank Ventures Andreessen Horowitz OS Fund	Investment	2018	Drug development	£7,888,600	13%	£60,103,243	n/a	n/a
2	Nuritas	European Investment Bank	Grant	2018	Drug development	£23,665,800	28%	£84,076,699	n/a	n/a
3	Insilico Medicine	WuXi AppTec Deep Knowledge Ventures	Investment	2018	Drug development	£4,733,160	7%	£70,279,537	2,000,000	£35
4	Recursion Pharmaceuticals	Square 1 Bank Bill & Melinda Gates Foundation	Investment	2018	Drug development	£47,331,600	26%	£178,905,559	5,200,000	£34
5	Berg	n/a	n/a	n/a	Drug development	n/a	n/a	n/a	n/a	n/a
6	BenevolentAI Ltd.	Woodford Investment Management	Investment	2018	Drug development	£90,718,900	5%	£1,668,438,900	50,000,000,000	£0
7	Covance	LabCorp	Acquisition	2014	Drug development	£5,157,692,898	100%	£5,157,692,898	100,000	£51577
8	Covance	LabCorp	Acquisition	2014	Drug development	£5,157,692,898	100%	£5,157,692,898	30,000,000,000	£0
9	Evidera	Pharmaceutical Product Development Inc.	n/a	n/a	Drug development	n/a	n/a	n/a	n/a	n/a
10	Atomwise	Monsanto	Investment	2018	Drug development	£35,498,700	20%	£180,609,497	600,000,000	£0
11	Genetech	Roche	Investment	2009	Drug development	£36,918,648,000	100%	£36,918,648,000	N/A	n/a
12	Antidote Technologies Inc.	Merck Global Health Innovation LLC	Investment	2017	Drug development	£8,677,460	16%	£55,220,200	N/A	n/a
13	Innoplexus AG	HCS Beteiligungsgesellschaft mbH	Investment	2016	Drug development	n/a	n/a	n/a	n/a	n/a
14	NuMedii Inc.	Claremont Creek Ventures	Investment	2015	Drug development	£1,577,720	11%	£14,538,690	n/a	n/a
15	Pharnext SA	Truffle Capital SAS	Investment	2016	Drug repurposing	£24,375,774	31%	£79,201,544	n/a	n/a
16	Qrativ Inc.	Mayo Medical Ventures	Investment	2017	Drug development	£6,547,538	17%	£39,663,881	n/a	n/a
17	Zephyr Health Inc.	Anju Software Inc.	Acquisition	2018	Drug development	n/a	n/a	n/a	n/a	n/a
18	e-Therapeutics	Octopus Ventures	Investment	2007	Drug development	£1,049,184	4%	£29,448,144	n/a	n/a
19	ExScientia Ltd.	Evotec AG	Investment	2017	Drug development	£11,832,900	31%	£38,307,042	n/a	n/a
20	ExScientia Ltd.	GT Healthcare Capital Partners LP	Investment	2019	Drug development	£20,510,360	n/a	n/a	n/a	n/a

Appendix – supplement (continued)

Private companies w/assets that include access to drug development data

Indicative estimates of market value for NHS data sets

Market-based approach: private companies benchmarking
Analysis as of 01 January 2019

	Target	Buyer	Type of deal	Year	Domain	Deal value	Approx % trans acted	Post-money valuation or acquired price	Number of records	Implied EV or record
Private companies w/assets that include access to drug development data								[1]	[2]	
21	Healx Ltd.	Amadeus Capital Partners Ltd.	Investment	2018	Drug development	£7,888,600	21%	£37,825,837	n/a	n/a
22	Numerate Inc.	Foundation Capital	Investment	2014	Drug development	£6,310,880	15%	£42,661,549	n/a	n/a
23	Owkin Inc.	Cathay Innovation	Investment	2018	Drug development	£14,199,480	16%	£87,547,683	n/a	n/a
24	ReviveMed Inc.	Rivas Capital LLC	Investment	2018	Drug development	£1,183,290	13%	£8,843,121	n/a	n/a
25	Alector Inc.	OrbiMed Advisors LLC	Investment	2019	Drug development	£138,642,145	13%	£1,105,776,616	n/a	n/a
26	BioAge Labs Inc.	Felicis Ventures	Investment	2017	Drug development	£8,598,574	20%	£43,276,860	n/a	n/a
27	Causaly Inc.	Marathon Venture Capital	Investment	2018	Drug development	£788,860	25%	£3,131,774	n/a	n/a
28	GTN Ltd.	Octopus Ventures	Investment	2018	Drug development	£2,208,808	30%	£7,446,838	n/a	n/a
29	Prellis Biologics Inc.	SOSV LLC	Investment	2017	Drug development	£1,459,391	29%	£5,064,481	n/a	n/a
30	Spring Discovery Inc.	First Round Capital	Investment	2018	Drug development	£14,199,480	17%	£84,108,253	n/a	n/a
31	XtalPi Inc.	China Life Investment Hodling Co. Ltd	Investment	2018	Drug development	£36,287,560	22%	£165,810,483	n/a	n/a
32	Lantern Pharma Inc.	Bios Partners	Investment	2017	Drug development	£2,918,782	35%	£8,440,802	n/a	n/a
33	Resonant Therapeutics Inc.	Mercury Fund Ventures III LP	Investment	2017	Drug development	£1,609,274	17%	£9,229,662	n/a	n/a
34	Envisagenics Inc.	Notion Capital	Investment	2017	Drug development	£2,761,010	13%	£21,291,331	n/a	n/a
35	Notable Labs Inc.	Builders VC LLC	Investment	2017	Drug development	£7,888,600	14%	£57,886,547	n/a	n/a
36	Celsius Therapeutics Inc.	Third Rock Ventures LLC	Investment	2018	Drug development	£51,275,900	30%	£168,326,947	n/a	n/a
	First quartile							£25,369,738	2,800,000	£0
	Median							£57,886,547	302,600,000	£17
	Third quartile							£167,068,715	22,650,000,000	£35

Notes:
EV=Estimated value
[1] Source: SP Capital IQ, DowJones VentureSource.
[2] Source: EY research and/or company websites.

Appendix-Supplement (continued)

Private companies w/assets that include access to population health data

Indicative estimates of market value for NHS data sets

Market approach: private companies benchmarking
Analysis as of 01 January 2019

	Target	Buyer	Type of deal	Year	Domain	Deal value	Approx % trans acted	Post-money valuation or acquired price	Number of records	Implied EV or record
Private companies w/assets that include access to population health data								[1]	[2]	
1	Health Fidelity	UPMC	Investment	2015	Pop Health	£15,193,444	22%	£68,859,589	200,000,000	$0
2	Lumiata	Sandbox Industries, Khosla Ventures, Intel Capital	Investment	2018	Pop Health	£8,677,460	10%	£85,827,968	20,000,000	$4
3	Innovaccer	WestBridge Capital	Investment	2018	Pop Health	£27,610,100	20%	£136,275,565	10,000,000	$14
4	Ginger.io Inc	Kaiser Permanente Ventures	Investment	2014	Pop Health	£15,777,200	18%	£89,330,506	2,000,000	$45
5	HealthCrowd	Healthy Ventures	Investment	2018	Pop Health	£5,679,792	12%	£48,270,343	n/a	n/a
	First quartile							£68,859,589	8,000,000	£3
	Median							£85,827,968	15,000,000	£9
	Third quartile							£89,330,506	65,000,000	£21

Notes:
EV=Estimated value
[1] Source: SP Capital IQ, DowJones VentureSource.
[2] Source: EY research and/or company websites.

Appendix-Supplement (continued)

Private companies w/assets that include synergies between multiple data types

Indicative estimates of market value for NHS data sets

Market approach: private companies benchmarking
Analysis as of 01 January 2019

	Target	Buyer	Type of deal	Year	Domain	Deal value	Approx % trans acted	Post-money valuation or acquired price	Number of records	Implied EV or record
Private companies w/assets that include synergies between multiple data types								[1]	[2]	
1	DeCODE Genetics	Amgen	Acquisition	2012	Genomics	£327,376,900	100%	£327,376,900	140,000	£2,338
2	The Advisory Board	Optum	Acquisition	2017	General	£1,025,518,000	100%	£1,025,518,000	100,000	£10,255
3	Ayasdi	Kleiner Perkins Khosla Ventures	Investment	2015	General	£43,387,300	20%	£214,656,695	n/a	n/a
	First quartile								110,000	£4,318
	Median								120,000	£6,297
	Third quartile								130,000	£8,276

Notes:
EV=Estimated value
[1] Source: SP Capital IQ, DowJones VentureSource.
[2] Source: EY research and/or company websites.

Acknowledgments

Sincere thanks from EY teams to the following individuals who assisted in the creation of this report.

Pamela Spence, EY Global Health Sciences and Wellness Industry Leader, and **Rob Moody**, TAS Chief Innovation Officer and EY UK&I TAS Analytics Leader, provided strategic guidance and feedback on the narrative throughout its development.

Peter Arnold, EY UK&I TAS Economics Lead, **David Gudmunsen**, EY UK&I TAS Valuations Lead, **Natasha Hunerlach**, EY US TAS Valuations Lead, and **Chris Wayman**, EY UK&I TAS Business Modelling and Decision Analytics, all provided key input on concepts of the publication.

Rajiv Mehrotra, **Himanshu Chaudhry**, **Hannah Kalozdi** and **Wade Robbins**, under the guidance of **Natasha Hunerlach**, performed industry research, detailed fair market value pricing for health care data sets, and conducted analysis associated with estimating the value of NHS data set value that underpinned the market-based approach.

George Agathangelou, **Emily Hutchison**, **Nida Mahmud** and **Sabina Mamedova**, under the guidance of **Peter Arnold**, conducted health care research and analysis to underpin the income-based approach quantifying the economic benefits associated with unlocking insights from NHS patient records.

James Evans reviewed the publication for quality and consistency. **Dennis Ryan** created the memorable accompanying visuals.

Authors

Chris Wayman, PhD
EY UK&I Life Sciences & Health Care – Valuation, Economics & Modelling
cwayman@uk.ey.com
+44 20 7951 6469

Natasha Hunerlach, CFA, CFE, MRICS
EY Americas Life Sciences & Health Care – Valuation, Economics & Modelling
natasha.hunerlach@ey.com
+1 404 817 5942

Contacts

EY GLOBAL HEALTH SCIENCES AND WELLNESS INDUSTRY LEADER

Pamela Spence
pspence2@uk.ey.com
+44 207 951 3523

EY GLOBAL TRANSACTIONS ADVISORY SERVICES

Peter Behner
peter.behner@de.ey.com
+49 30 25471 12467

EY LIFE SCIENCE & HEALTH CARE VALUATIONS, ECOMONICS & MODELLING

Peter Arnold
parnold@uk.ey.com
+44 20 7951 5587

David Gudmunsen
dgudmunsen@uk.ey.com
+44 20 7951 1519

Natasha Hunerlach
natasha.hunerlach@ey.com
+14048175942

Rob Moody
rmoody@uk.ey.com
+44 20 7951 2515

Chris Wayman
cwayman@uk.ey.com
+44 20 7951 6469

EY | Assurance | Tax | Transactions | Advisory

About EY
EY is a global leader in assurance, tax, transaction and advisory services. The insights and quality services we deliver help build trust and confidence in the capital markets and in economies the world over. We develop outstanding leaders who team to deliver on our promises to all of our stakeholders. In so doing, we play a critical role in building a better working world for our people, for our clients and for our communities.

EY refers to the global organization, and may refer to one or more, of the member firms of Ernst & Young Global Limited, each of which is a separate legal entity. Ernst & Young Global Limited, a UK company limited by guarantee, does not provide services to clients. Information about how EY collects and uses personal data and a description of the rights individuals have under data protection legislation are available via ey.com/privacy. For more information about our organization, please visit ey.com.

About EY Health Sciences and Wellness
The rise of the empowered consumer, coupled with technology advancements and the emergence of digitally focused entrants, is changing every aspect of health and care delivery. To retain relevancy in today's digitally focused, data-infused ecosystem, all participants in health care today must rethink their business practices, including capital strategy, partnering and the creation of patient-centric operating models.

The EY Health Sciences and Wellness architecture brings together a worldwide network of more than 20,000 professionals to build data-centric approaches to customer engagement and improved outcomes. We help our clients deliver on their strategic goals; design optimized operating models; and form the right partnerships so they may thrive today and succeed in the health systems of tomorrow. We work across the ecosystem to understand the implications of today's trends, proactively finding solutions to business issues and to seize the upside of disruption in this transformative age.

EYG no. 003378-19Gbl
1906-3182693 US CSG
ED none

ey.com/lifesciences