



Mining for Healthcare Gold

The Value of Healthcare Data

BY DAVID LO, CFA, ASA, BOBBY VAN DUSEN, AM, AND NICK JANIGA, ASA

Table of Contents

INTRODUCTION	3
DIGITIZATION OF CLINICAL DATA	4
ARTIFICIAL INTELLIGENCE.....	5
PRIVACY	6
VALUATION APPROACHES	13
FMV CONSIDERATIONS	20
SUMMARY	20
CONTACT US	21
ABOUT OUR HEALTHCARE CAPABILITIES	22
ABOUT STOUT.....	23

PRESS INQUIRIES

Please contact Sarah Walsh at sleas@stout.com for additional information.

Access our Media Kit at <https://www.stout.com/en/about/media-kit>.

DISCLAIMER

Informational purposes only

This material is for informational purposes only; it does not take into account any specific individual or entity's facts and circumstances. The content does not constitute legal, financial, accounting, or other professional advice, and should not be relied upon by you or any third party, including to operate or promote your business, secure financing or capital in any form, obtain any regulatory or governmental approvals, or otherwise be used in connection with procuring services or other benefits from any entity. Stout assumes no liability for viewers' use or use by any third party of the information herein and please note, before making any decision or taking any action, you should consult with legal and professional advisers. All opinions expressed are those of the authors and do not necessarily reflect the views of Stout Risius Ross, LLC, Stout Advisors SA, Stout Bluepeak Asia Ltd, Stout GmbH, MB e Associati S.r.l., Stout Park Ltd or Stout Capital, LLC.

Introduction

Healthcare is undergoing a transformation due to abilities to collect, analyze, and apply data in delivering healthcare services. Healthcare data will become increasingly important as organizations learn to leverage it to improve outcomes, reduce costs, study populations, close gaps in care, address labor challenges, and develop new treatments. Accenture Research estimates that 70 percent of healthcare workers' tasks could be reinvented through technology augmentation or automation.¹ With years of data at their disposal, an increasing number of healthcare companies are developing and promoting their artificial intelligence (AI)-driven solutions. While the capabilities of AI systems vary greatly, generally all require data for development.

In this report, we examine how companies across the healthcare industry have been using data, as well as the fair market value considerations when entering into data sharing and data use arrangements. We also provide insight into the different valuation approaches that can be used to value healthcare data.

¹ Shah, MD, Tejash, et al., "Reinvent care delivery to solve clinical shortage," Accenture, March 1, 2023.



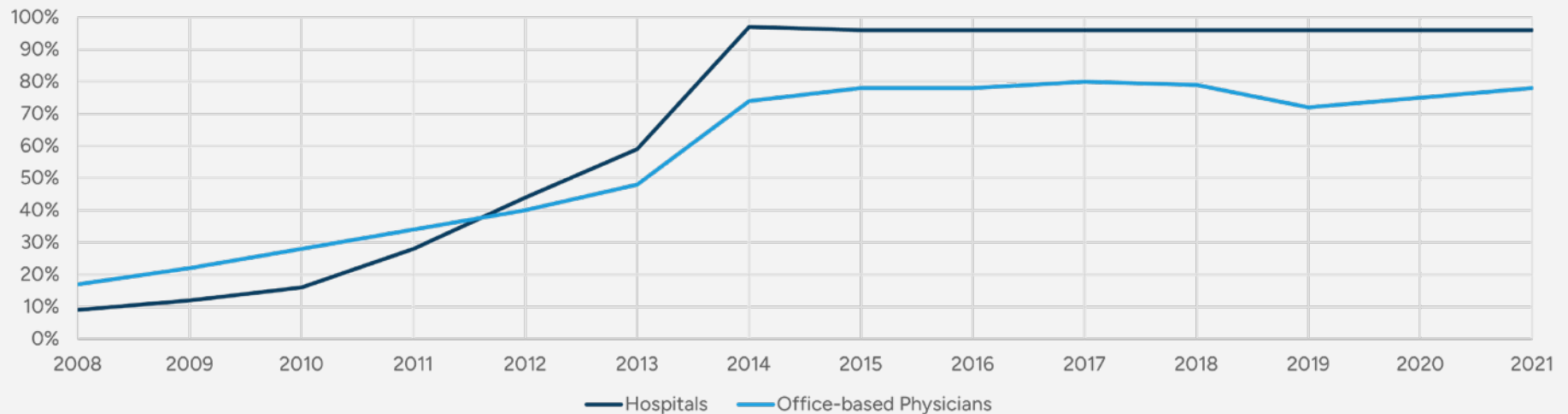
Digitization of Clinical Data

The history of electronic health records (EHRs) can be traced back to the 1960s when hospitals began experimenting with computerized systems to store patient information. However, widespread adoption gained momentum between 2008 and 2014 largely due to incentive programs implemented by the U.S. government. The Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009, part of the American Recovery and Reinvestment Act, allocated substantial funds to healthcare providers that demonstrated meaningful use of EHRs. This financial incentive, coupled with potential penalties for non-compliance, encouraged healthcare organizations to invest in and implement EHR systems, fostering a digital transformation in the healthcare sector. As

outlined in **Figure 1**, EHR adoption has now peaked at around 96 percent for hospitals and roughly 80 percent for office-based physicians.^{2,3,4}

This digitization of healthcare data was also seen in the diagnostic imaging space. Traditionally reliant on analog film, the transition to digital radiography began in the late 20th century with the adoption of computed radiography and, later, digital radiography. This shift offered numerous advantages, including faster image acquisition, immediate availability for viewing, and the ability to enhance and manipulate images for better diagnostic accuracy. Picture Archiving and Communication Systems (PACS) played a pivotal role in the digitization process by replacing traditional film-based storage with electronic storage and retrieval systems. PACS facilitated seamless image sharing, remote access,

Figure 1: EHR Adoption



2 "AHA Annual Survey Information Technology Supplement, 2008-Present," American Hospital Association Data & Insights.

3 "NCHS National Ambulatory Care Survey," National Center for Health Statistics.

4 "National Electronic Health Records Survey," Centers for Disease Control and Prevention.

and collaboration among healthcare professionals, fostering quicker and more informed decision-making.

The expanded use of electronic health records and advancements in imaging technology have contributed to the growth in healthcare data. The amount of healthcare data generated globally will continue growing exponentially, driven by factors such as an aging population, increased healthcare spending, and the increased popularity of wearable devices and sensors (e.g., smartwatches and fitness trackers).

While the amount of healthcare data that is generated each year is difficult to quantify, companies such as International Data Corporation and RBC Capital Markets estimate the total amount of healthcare data has exceeded thousands of exabytes (one exabyte is equal to 1,000,000 terabytes).^{5,6} This digitization of healthcare data fulfills the big data needs for companies to develop AI tools.

Artificial Intelligence

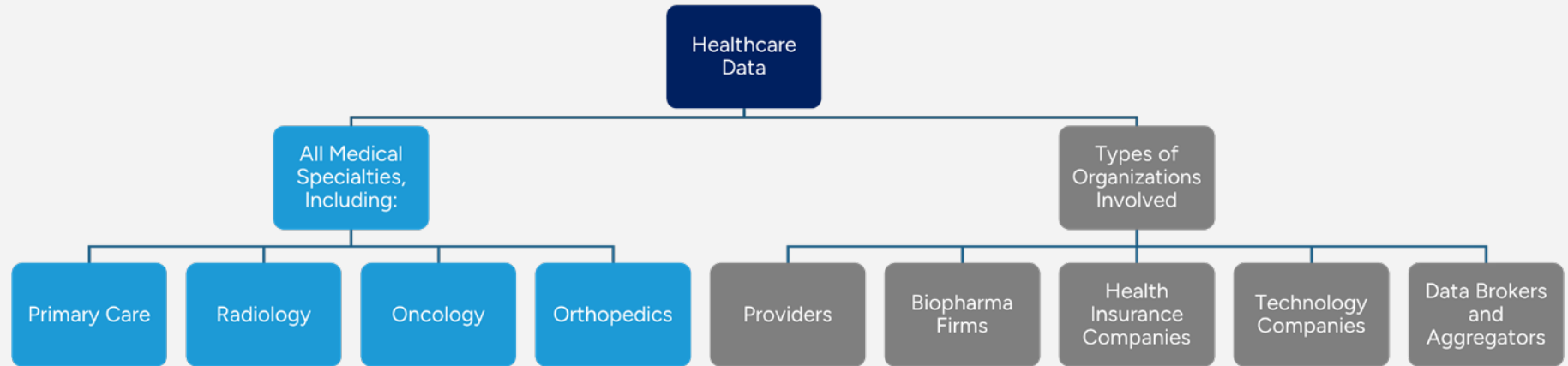
For decades, digital technology has been changing our daily lives. From email spam detection systems to virtual assistants, such as Siri and Alexa, our society interacts with AI on a daily basis. More recently, we have begun to experience the power of generative AI through applications such as Google Gemini and ChatGPT. As the capabilities of AI grow, we will continue to see companies develop AI solutions tailored to the healthcare industry.

Machine learning, a subset of artificial intelligence, focuses on creating algorithms and statistical models that enable computer systems to perform complex tasks without being explicitly programmed. One application of machine learning is structuring unstructured data found in medical notes, which can be achieved through natural language processing (NLP). Notably, an estimated 70% of the information in electronic health records (EHRs) is stored in free-text format. By applying NLP, this unstructured data can be structured, enhancing its usability and contributing to more effective healthcare research and outcomes.

5 "The Digital Universe, Healthcare Vertical Industry Brief," EMC Digital Universe.

6 Wiederrecht, Ph.D., Greg, et al., "The Healthcare Data Explosion," RBC Capital Markets.

Figure 2: Healthcare Data Key Stakeholders



Healthcare data is important for nearly all medical specialties and stakeholders across the continuum of care. It has been used to develop augmented intelligence, support research, advance precision medicine, and help shift healthcare from a fee-for-service model to a value-based care model.

Privacy

One unique aspect of data used within the healthcare industry is that personal health information is protected under the HIPAA Privacy Rule. Accordingly, this requires that the healthcare data be de-identified for use in many different applications. The failure of to comply with HIPAA can result in fines ranging from \$141 to \$71,162 per violation or per record, outlined in the following figure.⁷ In 2024, over \$9.1 million of fines or settlements were paid for HIPAA violations.⁸

⁷ "Annual Civil Monetary Penalties Inflation Adjustment," Health and Human Services Department, August 8, 2024.

⁸ "HIPAA Fines Listed by Year," Compliancy Group.

Figure 3: HIPAA Fines by Level of Culpability

Culpability	Minimum Penalty/ Violation	Maximum Penalty/ Violation	Annual Limit
No Knowledge	\$141	\$71,162	\$2,134,831
Reasonable Cause	\$1,424	\$71,162	\$2,134,831
Willful Neglect – Corrected within 30 Days	\$14,232	\$71,162	\$2,134,831
Willful Neglect – Not Corrected within 30 Days	\$71,162	\$2,134,831	\$2,134,831

*As of August 8, 2024^{9,10}

I. AUGMENTED INTELLIGENCE

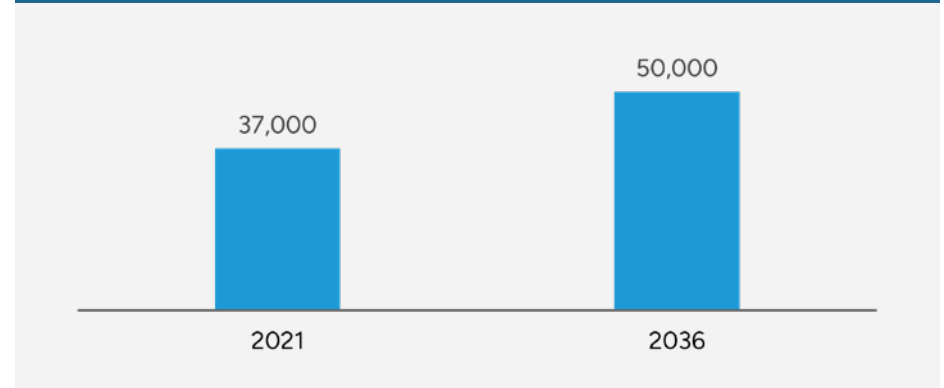
Healthcare providers are increasingly using augmented intelligence, a combination of artificial intelligence and human intelligence, to enhance patient care, operational efficiency, and decision making.

Clinical Decision Support

Many companies have developed decision-support tools to aid clinical decision makers due to a growing shortage of physicians. This shortage is driven by an increased need for healthcare services as the United States population ages. According to data published by the Association of American Medical Colleges (AAMC), the United States population is forecasted to grow 8.4 percent through 2036, while the population of adults aged 65 and older is expected to grow by 34.1 percent. Additionally, as of 2021, physicians aged 54 years and older

represent 42 percent of the physician population in the United States. The following figure highlights data published by AAMC on the current physician shortage in the United States and the midpoint of their forecasted range of physician shortage in 2036.

Figure 4: Physician Shortage in the United States



This macro trend also impacts radiologists, as the number of imaging studies is increasing by 5.0 percent per year, while the number of radiology residency positions is only increasing by approximately 2.0 percent per year.¹¹ In response to these shortages, AI-powered medical diagnostics can help reduce gaps in care, expedite decisions, and facilitate timelier delivery of care. Companies such as Tempus Radiology (f.k.a. Arterys) and Holberg EEG have developed tools that are able to interpret radiology images or clinical electroencephalograms (EEGs). Universities such as Northwestern Medicine have also developed their own in-house tools, using their clinical data to create a tailored solution. While these tools are not used on a standalone basis, they augment physicians by aiding in the detection and interpretation of clinical data and, with generative AI, produce reports for physicians to review and finalize. The development of these systems requires the use of structured training data that has been appropriately

9 While adjustments to the civil penalties are to be made by January 15 of each year, they can end up being delayed (e.g., the inflation adjustment in 2024 was not made until August). The cost-of-living adjustment multiplier for the 2025 adjustment is 1.02598.

10 "Implementation of the Federal Civil Penalties Inflation Adjustment Act and Adjustment of Amounts for 2025," National Aeronautics and Space Administration, May 7, 2025

11 Asin, Stefanie, "The Radiologist Shortage, Explained," Becker's Hospital Review, December 31, 2024.

annotated so that an AI system can “learn” from these examples. In a major clinical study, an AI tool boosted productivity by up to 40 percent without compromising accuracy.¹²

In other examples, studies have leveraged machine learning to identify patients at high risk of experiencing acute clinical needs or hospitalization using EHR-sourced data. These patients were then targeted for supplemental evaluation and preventative intervention. This ultimately results in a reduction in long-term costs.¹³ Ontrak (OTCPK: OTRK) uses AI to identify members living with unaddressed behavioral health needs whose clinical needs are likely to relate to undiagnosed conditions. Studies have shown that people with untreated behavioral health problems have higher overall healthcare costs, especially related to other chronic health conditions such as hypertension and diabetes. Ontrak deploys an AI algorithm to scan millions of data points related to historical claims data and, with a reportedly high degree of accuracy, infer behavioral health diagnoses.¹⁴

Predictive Analytics

Healthcare companies such as Medtronic (NYSE: MDT) and Alphatec Spine (NASDAQ: ATEC) have developed surgical planning software with predictive modeling capabilities, enabling surgeons to more precisely determine the surgical approach with the highest probability of success. This is done in an iterative cycle in which a surgeon generates an alignment report, simulates surgical scenarios, uses EMR data and patient-specific implants optimized for the surgical plan, and reconciles the surgical results with the pre-operative plan. With each procedure, post-operative data is collected and analyzed to refine predictive modeling for future surgeries.

The development of AI requires data, the quality of which directly impacts its value and the capability of the resulting AI systems. Accordingly, such data is highly valuable, but determining its value can take many forms.

II. RESEARCH / BIOPHARMA / CLINICAL TRIALS

Precision Medicine

The 21st Century Cures Act, signed into law in December 2016, aimed to accelerate the discovery, development, and delivery of new medical treatments and technologies. One of its key provisions focused on increasing the use of real-world evidence (RWE) in regulatory decision-making processes. RWE provides insights into how medical products perform in real-world settings and complements data from traditional randomized controlled trials. To derive RWE, real-world data (RWD) must be gathered and analyzed. RWD refers to data collected from sources outside traditional clinical trials, such as electronic health records, claims data, and wearable devices.

Biopharmaceutical companies have been utilizing RWE and RWD to spearhead the development of precision medicines, ushering in a new era of tailored therapeutics. One application of clinico-genomic data in biopharma is the identification of biomarkers associated with specific diseases or patient populations, a process performed using artificial intelligence. By analyzing genetic variations and correlating them with clinical outcomes, companies can pinpoint biomarkers indicative of disease risk, progression, or therapeutic response. This knowledge enables the development of targeted therapies designed to address the underlying genetic mechanisms driving disease.

The use of clinico-genomic data spans the entire drug development process, as outlined in **Figure 5**.

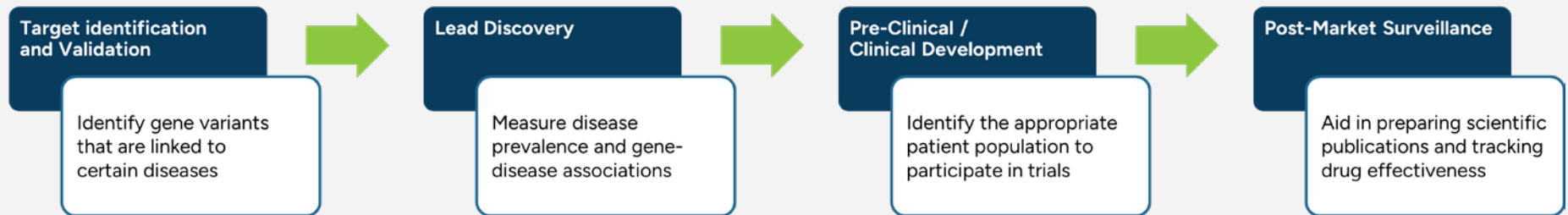
Clinico-genomic data holds immense value, but these databases are difficult to create and are often formed through partnerships among various entities. For example:

¹² Schamisso, Ben, “New AI Transforms Radiology with Speed, Accuracy Never Seen Before,” Northwestern University, May 29, 2025.

¹³ Natesan, M.D., et al., “Health Care Cost Reductions with Machine Learning—Directed Evaluations during Radiation Therapy — An Economic Analysis of a Randomized Controlled Study,” New England Journal of Medicine, March 15, 2024.

¹⁴ “How AI Technologies are Transforming the Member Journey in Behavioral Health,” Ontrak Health, November 8, 2022.

Figure 5: Data Usage Throughout Drug Development Process



- Biopharmaceutical firms have collaborated in the past to create clinico-genomic databases. One example is the American Association for Cancer Research's Project Genomics Evidence Neoplasia Information Exchange (Project GENIE). The goal of Project GENIE was to gather clinical and genomic data from approximately 50,000 patients. This data would then be utilized to make advances in precision oncology and clinical decision making. To achieve this goal, nine (9) of the largest biopharmaceutical firms contributed a combined \$36 million toward Project GENIE.¹⁵ In exchange, the biopharmaceutical companies gained exclusive access to the data for a limited time before it was released to the public.
- Flatiron Health and Foundation Medicine created a clinical-genomic database by combing the genomic profiling data sequenced with Foundation Medicine's assays with longitudinal clinical data developed by Flatiron Health through electronic health records.¹⁶
- Helix is a population genomics company that has partnered with major health systems across the U.S. to help them better understand the populations they serve. By providing no-cost testing, health systems have been able to recruit and provide genomic testing to large patient

populations. This data can be used by Helix and the health systems to develop genetic insights regarding patient risks that health systems are able to proactively manage.

- In other instances, companies such as IQVIA have purchased data from numerous sources, including biobanks, healthcare systems, research studies, recruited registries, healthcare providers, and pharmacies. Through this process, they aggregate and resell this information once sufficient scale has been achieved.
- While some physician practices may be large enough to develop their own databases, there has been a trend where practices partner with or affiliate with physician practice management companies (PPMCs). These PPMCs are able to help coordinate resources and develop databases through utilizing data from affiliated physician practices.

Companies that hold de-identified data assets are able to utilize the data themselves or commercialize it through data sharing or data use arrangements. The following figure highlights a few companies that hold RWD assets.

15 "AACR Project GENIE Begins Five-Year Collaborative Research Project with \$36 Million in New Funding," American Association for Cancer Research, October 31, 2019.

16 Snow, Tamara, "Clinico-Genomic Data is a Game Changer for Precision Oncology," Flatiron Health, March 2023.

Figure 6: Companies with Large Real World Data Assets

Company	Dataset Description
Flatiron Health	Health records for over 5 million patients in four of the largest oncology markets in the world. Included in the dataset is a clinical-molecular database of over 78,000 linked patients and a clinico-genomic database that combines clinical data and genomic profiling data for over 100,000 patients.
Helix	Dataset of over 1 million clinico-genomic records from a patient population across the United States. The dataset includes Whole Exome+ sequencing and longitudinal clinical data that is updated regularly. The dataset is currently being expanded through direct recruitment and enrollment from the source EHR system.
IQVIA	IQVIA holds a data asset that contains health information from over 1.2 billion patients. The data asset includes various subsets of data including linked clinico-genomic records and oncology treatment data.
Veradigm	Dataset containing information sourced from EHRs of over 154 million unique patients. Veradigm utilizes its proprietary natural language processing (NLP) models to extract meaningful elements from unstructured or semi-structured data.
Optum / UnitedHealth Group	Dataset of over 300 million patients that includes EHRs, commercial claims data, socioeconomic status data, and Medicare fee for service claims data.
Carelon Research / Elevance Health	Dataset of over 90 million patients that includes EHR, clinical oncology data, lab results, social determinants of health, and demographic data.
Labcorp	Dataset that contains lab results of over 150 million patients. The data asset is mainly utilized by licensees in clinical drug development.
Quest Diagnostics	Dataset that contains over 70 billion lab test results across virtually all therapeutic areas.

Source: Publicly available information from each company's website (as listed below).

- <https://flatiron.com/real-world-evidence/multimodal-data>
- <https://www.helix.com/life-sciences>
- <https://iqvia.com/solutions/real-world-evidence/real-world-data-and-insights>
- <https://team.veradigm.com/nlp-real-world-data-solutions>
- <https://labs.optum.com/>
- <https://www.carelonresearch.com/carelon-real-world-data>
- <https://labcorp.com/biopharma/commercialization/data-tech>
- <https://www.questdiagnostics.com/business-solutions/life-sciences/pharma-solutions/data-licensing>

III. VALUE-BASED CARE

Value-based care is a rapidly growing healthcare service model that prioritizes the quality of care over the quantity of services provided. Since physician compensation may be tied to the quality of care provided, the ability to measure and track the quality of care requires healthcare data. With the increased push to adopt value-based payment models in the United States, patient reported outcome measures (PROM) and patient reported outcome (PRO) data enable consideration of the patient perspective when measuring care quality.

Patient Reported Outcome Data

PROM and PRO data are now regularly utilized by providers. The National Library of Medicine has five categories of PROM, which include 1) health-related quality of life (HRQL); 2) functional status; 3) symptoms and symptom burden; 4) health behaviors; and 5) the patient's healthcare experience.

The Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip Arthroplasty and/or Total Knee Arthroplasty Performance Measure¹⁷ (IQR THA/TKA PRO-PM) was implemented as part of CMS' 2023 Inpatient Prospective Payment System (IPPS) Final Rule. The goal of the IQR THA/TKA PRO-PM is to measure pain and functional improvements in patients through gathering PRO data. The IQR THA/TKA PRO-PM requires Medicare-licensed hospitals to gather the necessary PRO data for 50.0 percent of their Medicare fee-for-service patients who undergo elective total hip or total knee arthroplasty procedures in an inpatient setting.^{18, 19}

The PRO data can be gathered from patients before, during, or after care. The data can be collected through various avenues, including emails, phone calls, online portals or surveys, direct mailers, or in person visits. The following figures are from CMS and shows examples of when PRO preoperative and postoperative data may be gathered:

17 This measure is also known as the CMS Inpatient Quality Reporting THA/TKA Patient-reported Outcomes Performance Measure.

18 PRO data gathering requirements began for all procedures on or after July 1, 2024.

19 "The Mandatory Centers for Medicare & Medicaid Services Inpatient Quality Reporting Total Hip Arthroplasty/Total Knee Arthroplasty Patient-reported Outcomes Performance Measure," American Academy of Orthopaedic Surgeons.



Figure 7: CMS Preoperative PROM Examples²⁰

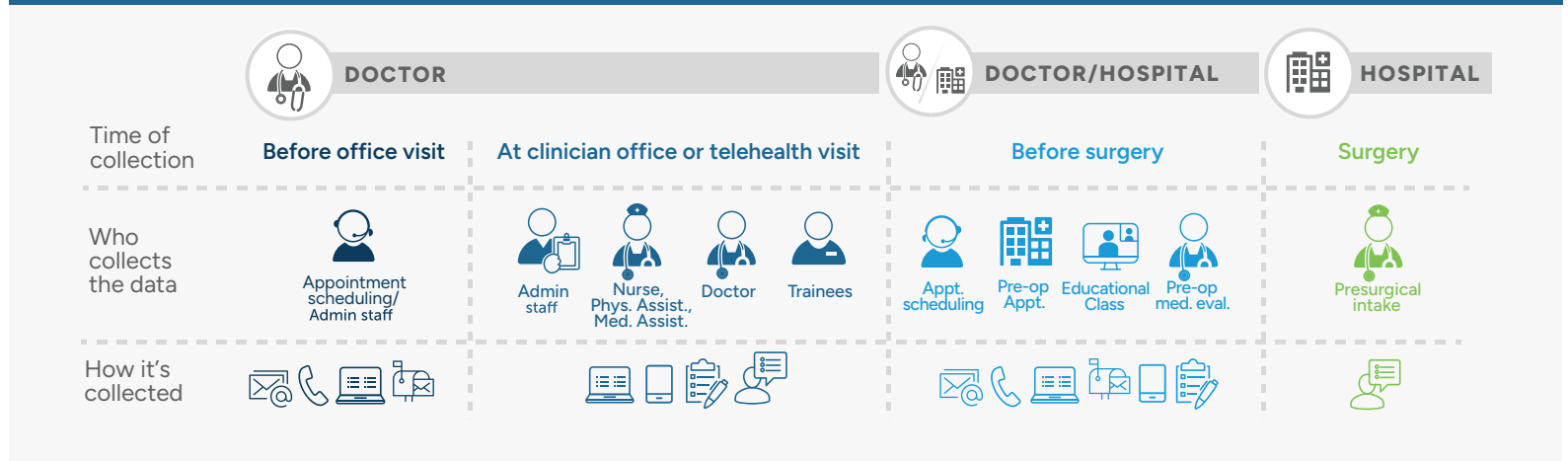
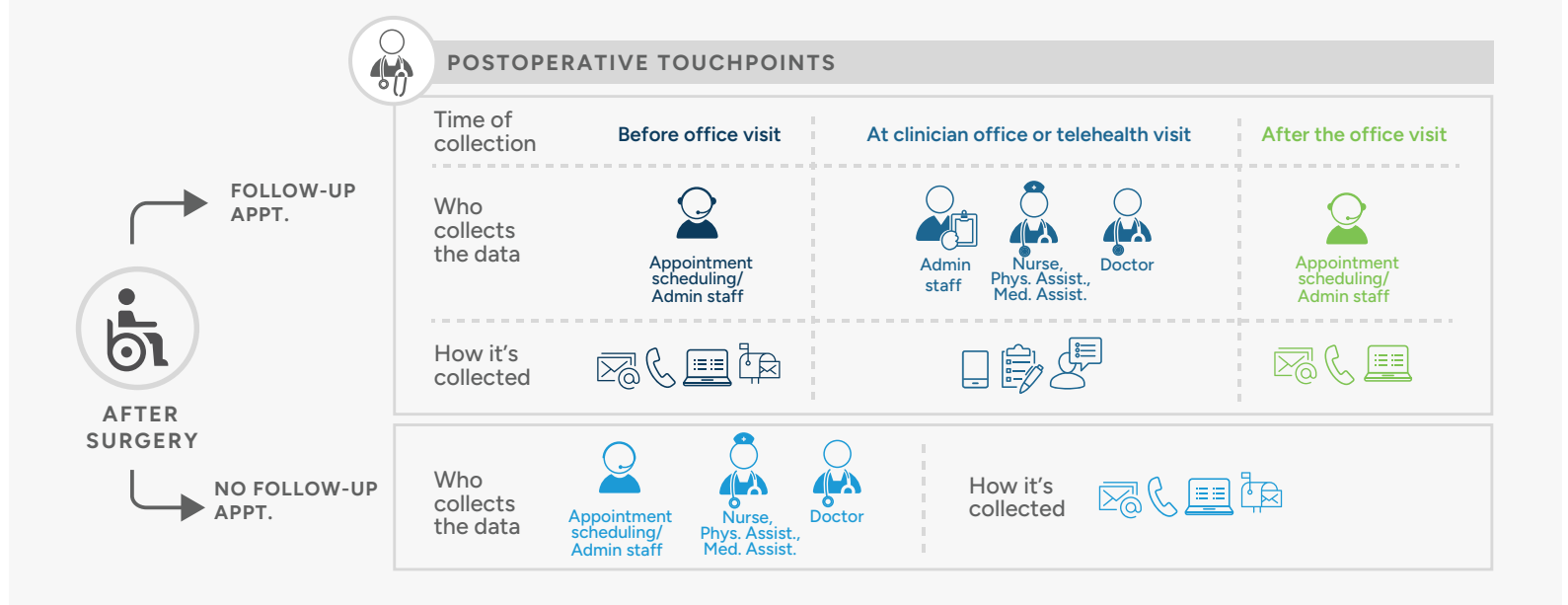


Figure 8: CMS Postoperative PROM Examples²¹



²⁰ "How and When can Patient-Reported Outcome (PRO) Data be Collected?" Centers for Medicare & Medicaid Services.

²¹ "How and When can Patient-Reported Outcome (PRO) Data be Collected?," Centers for Medicare & Medicaid Services.

CMS indicated that if hospitals fail to report a complete PRO dataset for 50% of eligible patients, a 25% reduction in the annual payment update will apply to all the hospital's fee-for-service Part A claims. CMS estimates that its annual payment updates typically range from 2.0 percent to 4.0 percent. Therefore, a hospital that fails to meet the reporting requirements could see a 0.5 percent to 1.0 percent reduction in its Medicare fee-for-service revenue.²²

From our experience, it is common for the hospital required to report the data to lack patient touchpoints to gather postoperative data. Therefore, they rely on physician practices for this information. A recent study evaluating PRO data collection efforts related to the IQR THA/TKA PRO-PM found that significant resources and a digital care platform was necessary to meet the reporting requirements.²³ Accordingly, we have observed arrangements whereby health systems may compensate a physician practice for this data. While the current use of this data is to ensure hospitals comply with reporting requirements, the use of PRO data will be valuable as providers transition to value-based care.

Valuation Approaches

The valuation of data, as for most intangible assets, relies on three established valuation approaches: the Market, Income, and Cost Approaches.

A. MARKET APPROACH

The application of the Market Approach requires the identification of comparable transactions where similar types of data sets have been utilized in arrangements and/or transacted. Healthcare data is commonly exchanged among organizations across the entire healthcare industry. These arrangements and transactions are typically private, and comparable information is generally not readily available. Stout sources comparable data from its proprietary database, research findings, and surveys of industry executives with direct knowledge of data asset transaction prices or arrangement terms. In our experience, the value of healthcare data can range from a fraction of a penny to thousands of dollars per record depending on the type of data.

Healthcare Data Related Transactions in the Marketplace

A. 23andMe

One example of a transaction involving genomic data involves 23andMe. 23andMe was the first company to receive approval from the Food and Drug Administration to offer direct-to-consumer genetic tests. After processing the tests in its labs, 23andMe provided its customers with reports that detail their genomic makeup. 23andMe also licensed the data it gathers to pharmaceutical and biotechnology firms for research, with a recent SEC filing stating it was under a licensing arrangement with GlaxoSmithKline.

In 2021, 23andMe went public through the SPAC, VG Acquisition Corp.²⁴ After going public, 23andMe operated as 23andMe Holding Co. 23andMe struggled with a lack of recurring customers and faced various data security and privacy concerns related to the use of the genomic data it gathered. In October of 2023, a data breach that affected approximately 7 million customers was disclosed by 23andMe. Following the data breach, 23andMe faced several lawsuits alleging that the company violated privacy laws. In September of 2024, seven independent directors resigned, and in March of 2025, the company declared bankruptcy.

With the bankruptcy announcement, 23andMe's CEO immediately resigned, stating she wanted to be "...in the best position to pursue the company as an independent bidder." In the following weeks, an auction began for the assets of 23andMe, with the pharmaceutical firm Regeneron winning the auction. Regeneron's winning bid was \$256 million for substantially all of 23andMe's assets, including its collection of genomic records. Regeneron leadership announced as part of the auction win that its goal was to continue "to use large-scale genetics research to improve the way society treats and prevents illness overall." However, in June of 2025, 28 attorney generals sued 23andMe and blocked the auction to protect the genomic data. After the first auction was blocked, TTAM Research Institute, a non-profit company owned by 23andMe's former CEO, won the second auction with a bid of \$305 million. A federal

²² "Mandatory CMS Inpatient THA/TKA PRO-PM Frequently Asked Questions," American Academy of Orthopaedic Surgeons.

²³ Ghoshal, Soham, et al., "Evaluation Patient-Reported Outcome Measure Collection and Attainment of Substantial Clinical Benefit in Total Joint Arthroplasty Patients," The Journal of Arthroplasty, November 23, 2024.

²⁴ Special Purpose Acquisition Company.



bankruptcy court later approved the transaction. In a recent article from The HIPAA Journal, the purchase price includes the genomic data of approximately 13 million customers, plus the genome service and research services business lines and 23andMe's Lemonaid telehealth business.²⁵

Given the publicity surrounding this bankruptcy transaction, the purchase price of the assets and the number of 23andMe customers can be used to estimate the purchase price of the genomic data. However, consideration must first be given to the value of other assets transacted as part of the \$305 million purchase price. Based on information gathered from 23andMe's SEC filings and S&P Capital IQ, we removed value attributable to 23andMe's fixed assets and identified intangible assets (utilizing reported financial statements prior to the bankruptcy announcement). After adjusting the purchase consideration, we estimate that the value of 23andMe's genomic data is approximately \$20 per customer. It is worth noting that this estimated price may be discounted since the transaction occurred as a result of a bankruptcy filing.

B. Invitae

Another example of a transaction involving genomic data involves Invitae. Invitae offers genetic tests in a variety of clinical areas, including oncology, women's health, and rare diseases. In December of 2022, Invitae released a data transparency report, which detailed how it used the data collected from genetic tests. The report highlighted that Invitae's de-identified data contributed to 38 peer-reviewed publications, 36 posters, and 21 abstracts. Additionally, Invitae participated in a program where patients could elect to take free genetic tests, with the de-identified data being offered to biopharma firms for the development of new treatments. The report stated that Invitae had a database of over 2 million tested individuals at the time of the report.

In 2015, Invitae went public, with its stock price rising to an all-time high in 2020. The following year, it acquired a startup company focused on enabling customers to access and organize their health records. Costs began to climb for Invitae after the acquisition. It divested the startup company and performed two rounds of staff layoffs to combat the rising costs. In 2024, Invitae was delisted from the New York Stock Exchange and declared bankruptcy in February.

In August of 2024, Labcorp announced that it acquired select assets of Invitae for \$239 million. Labcorp offers comprehensive laboratory services that aid doctors, hospitals, pharmaceutical companies, researchers, and patients to make informed medical decisions. As Labcorp is also publicly traded, the purchase price allocation related to the acquisition of select assets of Invitae was published in Labcorp's 10-K filing. We note that Labcorp's 10-K states that the identified intangible assets include non-compete agreements, customer relationships, trade names, and technology assets. As Invitae's genomic data is not included in this asset class, the value of the data would be represented by all or part of the residual goodwill value. Using the goodwill value from the purchase price allocation as well as publicly available data that Invitae's database contained over 2 million records, we have estimated that the value per record for Invitae's genomic data ranges up to \$50. The following table summarizes the purchase price allocation as well as our estimate of the value of the genomic data included in the transaction:

²⁵ The HIPAA Journal states that 23andMe has 15 million customers but approximately 2 million of the 15 million customers elected to have their genetic data and biological samples destroyed.

Figure 9: Estimated Value of Invitae's Genomic Data

Asset	Allocated Value	
Inventory	\$12,100,000	
Property, Plant, and Equipment	76,700,000	
Goodwill	100,400,000	
Identifiable Intangible Assets	113,200,000	
Total Assets ^(a)	\$302,400,000	
Genomic Data Value Estimate	Low	High
Goodwill	\$100,400,000	
% Attributable to Genomic Data	50%	100%
Estimated Value of Genomic Data	50,200,000	100,400,000
Number of Records	2,000,000	2,000,000
Estimated Value per Record	\$25	\$50

Source: LabCorp 2024 10-K and Stout estimates

(a) LabCorp also acquired \$61.6 million in liabilities which reduced the cash consideration to \$240.8 million.

As with the 23andMe transaction, this value may be understated given the asset was acquired as a result of Invitae filing for bankruptcy. In comparing the estimated value of the data for 23andMe and Invitae, we note that Invitae's genomic tests are typically ordered by providers for their patients, while 23andMe tests are direct-to-consumer without any provider interaction. Because a provider



was typically involved with Invitae's genomic testing, the results may be integrated with other healthcare data and include review and notes from a genetic counselor. In contrast, 23andMe test results likely have limited (or no) connection to other healthcare data and have limited (or no) review notes by a provider.

Similar to the value estimates for 23andMe and Invitae, we have reviewed transactions involving data assets from other companies to assess similar estimates. The following figure highlights broad ranges of value we have observed and estimated for certain types of healthcare data assets:

Figure 10: Estimated Ranges of Value by Type of Data

Electronic Health Records		Imaging Data		Clinico-Genomic Data	
Low	High	Low	High	Low	High
\$2	\$50	\$0.01	\$450	\$200	\$2,000

Source: Publicly available information as well as Stout observations and estimates.

We note that the contents and quality of a subject data asset, as well as the terms of the contemplated transaction, significantly influence value and that these ranges are for informational purposes only.

In addition to the observed range of value for healthcare data, we have also observed various datasets that are free to the public for research-related purposes. A few examples of freely accessible healthcare databases are highlighted in **Figure 11**.

Market Approach Considerations

Valuing healthcare data can be complex and highly context dependent. Several general considerations when valuing healthcare data are outlined in **Figure 12**.

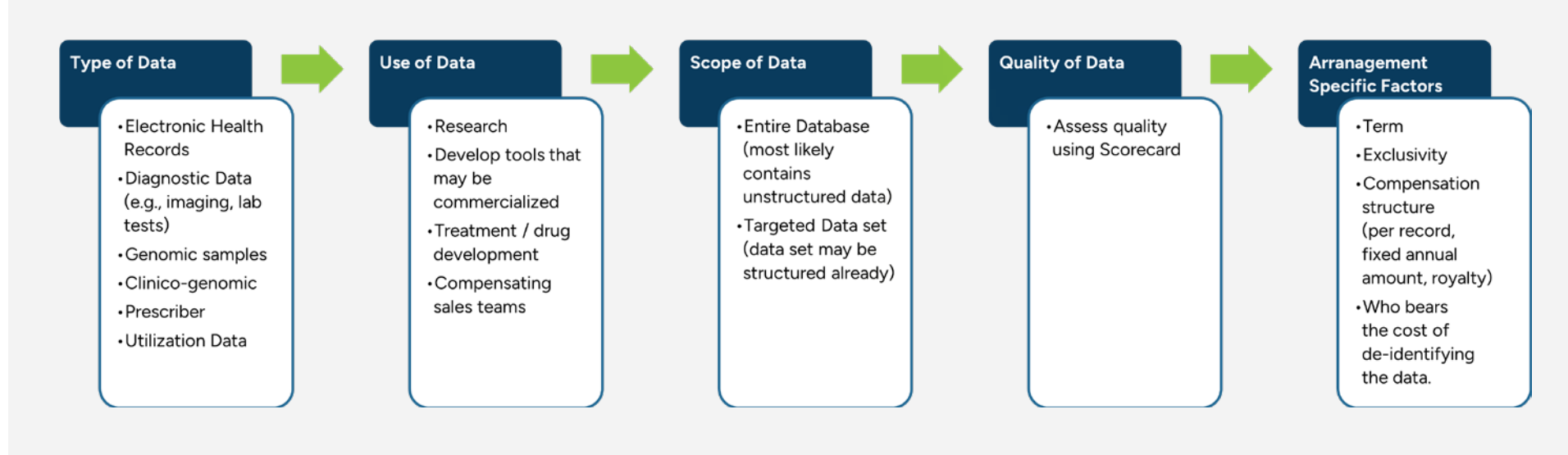
To determine the value for data assets being valued, Stout compiles a list of comparable transactions for similar types of data (e.g., radiology data, EHR, clinico-genomic data). After identifying comparable values, we use a scorecard to assess the “quality” of the subject data as summarized in **Figure 13**.

Figure 11: Examples of Free to Access Healthcare Data

Dataset	Description
AACR Project GENIE	Publicly accessible cancer registry of real-world clinico-genomic data gathered from data sharing arrangements with 20 international cancer centers.
Stanford AIMI 100 Genomes	The largest fully open resource of whole-genome sequencing (WGS) data that is available to the public without access or use restrictions.
USCF Clinical Data	A set of de-identified clinical and dental care data for approximately 4.3 million patients that is available for research use.

Source: Publicly available information from the organizations’ websites.

Figure 12: General Data Valuation Considerations



A further description of the healthcare data quality factors are outlined below:

1. **Completeness:** This refers to the amount of data contained within a patient’s EHR. Patients who have medical records that contain both primary care and episodes of secondary care will generally have a greater amount of data that may be useful for future research. Due to the COVID-19 pandemic, we have observed the volume of patient records increase significantly since 2020; however, a subset of these records often contain small amounts of useful clinical data. This is an example of an EHR that is incomplete.
2. **Accuracy and Plausibility:** This refers to whether the data contained within an EHR is accurate. Data that must be manually entered is susceptible to human error.
3. **Currency:** This refers to whether the data is up to date. A study was conducted on how varying sets of longitudinal historical training data can impact the prediction of future clinical decisions. As clinical practice patterns can vary across years, it was shown that small amounts of recent data are more effective than using larger amounts of older data toward future clinical predictions.²⁶ Accordingly, we believe that when assessing the value of clinical data, it’s important to consider whether the records being licensed represents active or inactive patients and how current the data is. While the definition of active and inactive patients can vary by organization and/or clinical specialty, we commonly see this defined as patients seen within the last 24 to 36 months.
4. **Relevance:** Depending on the type of clinical data, the relevance of the data may decline over time as there are changes in clinical protocols. This may not always be the case, as certain rare disease data remains relevant due to the lack of data, making it highly valuable.

Healthcare Data Quality Factors	Description	Weighted Score
1. Completeness	This refers to the amount of data contained within a patient's EHR.	
2. Accuracy / Plausibility	This refers to whether the data contained within an EHR is accurate.	
3. Currency	This refers to whether the data is up to date.	
4. Relevance	This refers to how well the data applies to a specific task or research area.	
5. Conformity	This refers to whether the data is contained in a standardized format or if it is unstructured.	
6. Bias	This refers to whether the data is representative of the broader population.	
7. Regulatory / Security	This refers to whether the data has been de-identified according to regulatory requirements.	
8. Provenance	This refers to the documentation of a piece of data's origin, history, custody, and transformations.	
Sample Weighted Score (Out of 5)		3.63

²⁶ Chen, Jonathan, et al., "Decaying Relevance of Clinical Data Towards Future Decisions in Data-Driven Inpatient Clinical Order Sets," International Journal of Medical Informatics, March 18, 2017.

5. **Conformity:** This refers to whether the data is contained in a standardized format or if it is unstructured. Most electronic health records contain unstructured data that is not very useable until it has been structured. With large language models, AI is able to assist in structuring this data. Historically, manual chart review has had to be performed, which is a time intensive process. Accordingly, data that has been structured commands a premium value.
6. **Bias:** Bias in data can occur when the data is only representative of certain subsets of the population as opposed to the broader population. Insights that may be developed from biased samples may not be broadly applicable. One paper indicated that there may be a pattern where sicker patients have higher levels of data completeness, which implies that exclusion based on complete records will select a biased sample in terms of patient health levels.²⁷
7. **Regulatory and Security:** In the United States, the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule restricts the transfer of health information. Specifically, HIPAA provides two standards for the disclosure of health information without seeking patient authorization: (1) Safe Harbor and (2) Expert Determination. The Safe Harbor method requires the removal of 18 types of identifiers.²⁸ The Expert Determination method uses statistical and scientific principles and methods to render the information not individually identifiable. The methodology used to de-identify data impacts the usefulness of the data and whether the de-identified data can be tokenized such that longitudinal insights can still be derived from the de-identified data.

When developing new treatment protocols or tracking the efficacy of treatments and drugs, the ability to track an individual patient over time becomes important. This is why we also see that there is a premium attached to longitudinal data as compared to cross-sectional data.

8. **Provenance:** Provenance refers to the documentation of a piece of data's origin, history, custody, and transformations. As databases get combined and shared, the ability to audit the underlying data ensures its integrity.

Exclusivity

When determining the value of healthcare data, it's also important to consider whether the licensing arrangement will be on an exclusive or non-exclusive basis, and whether there are restrictions on the potential use. As an exclusive arrangement would limit the number of potential licensees, a licensor would charge a premium when entering into an exclusive arrangement.

In 2018, 23andMe and GSK entered into a collaboration agreement. In consideration for 23andMe granting GSK exclusive access to the 23andMe database and mining technologies, GSK paid a total of \$150,000,000 over a five-year period, which represented an annual average of \$30,000,000.²⁹ Subsequently, 23andMe announced a new non-exclusive data licensing agreement with GSK that included a \$20,000,000 upfront payment, which represents approximately a 33% discount.³⁰ While exclusivity provisions are not the only differences in the two agreements, this provides some high-level insights into potential pricing differences between exclusive and non-exclusive arrangements.

All healthcare organizations have data. The raw value of the data may decline

²⁷ Weiskopf, Nicole, et al., "Sick Patients Have More Data: The Non-Random Completeness of Electronic Health Records," AMIA Annual Symposium Proceedings, November 16, 2013.

²⁸ The 18 identifiers include, 1) Names; 2) All geographical subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geocodes, except for the initial three digits of a zip code, if according to the current publicly available data from the Bureau of the Census: (1) The geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and (2) The initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000; 3) All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older; 4) Phone numbers; 5) Fax numbers; 6) Electronic mail addresses; 7) Social Security numbers; 8) Medical record numbers; 9) Health plan beneficiary numbers; 10) Account numbers; 11) Certificate/license numbers; 12) Vehicle identifiers and serial numbers, including license plate numbers; 13) Device identifiers and serial numbers; 14) Web Universal Resource Locators (URLs); 15) Internet Protocol (IP) address numbers; 16) Biometric identifiers, including finger and voice prints; 17) Full face photographic images and any comparable images; and 18) Any other unique identifying number, characteristic, or code (note this does not mean the unique code assigned by the investigator to code the data).

²⁹ According to SEC filings, GSK paid \$25,000,000 per year for the initial 4 years of the contract and had the right at its discretion to extend the term of the agreement by 1 year or an additional \$50,000,000.

³⁰ "23andMe Announces Collaboration Extension with a New Data Licensing Agreement with GSK," 23andMe, October 30, 2023.

as the “supply” of it continues to increase. The supply of data available to researchers is increasing rapidly, as evidenced by the availability of free-to-access healthcare data exemplified earlier in this article. While de-identified data can be commercialized as is, its potential value can be significant as organizations find ways to develop tools and insights from it. In cases where organizations have been able to commercialize the data, or develop tools using the data, an Income Approach may be applicable.

B. INCOME APPROACH

Under the Income Approach, a projected stream of economic benefits attributable to the healthcare data is discounted to a present value using a risk-adjusted rate of return. Although several methodologies are categorized under the Income Approach, a common methodology is the multi-period excess earnings method (MPEEM). Under this method, the economic benefits attributable to the business, service line, or product line are considered the sum of returns from the individual assets of the business, service line, or product line—one of which is presumed to be healthcare data.

The process begins with projecting the business’ overall economic benefits. Next, the stream of benefits is allocated among the various assets owned and utilized by the business, including working capital, fixed assets, and intangible assets (e.g., licenses, technology, workforce, etc.). A risk-appropriate rate of return and value are set for each asset. These inputs allow for the calculation of a contributory charge for each of the identified assets. The contributory charge for each identified asset is subtracted from the total economic benefit of the entire business; the resulting economic stream is deemed to be attributable to the healthcare data. The economic stream attributable to the data is discounted at a risk-adjusted rate of return to determine the present value of the healthcare data.

C. COST APPROACH

The Cost Approach determines the value of an asset based on the cost to create or re-create the asset. In the case of large, longitudinal clinico-genomic data sets, it may be unreasonable for an organization to try to replicate the database because this data was originally gathered as a byproduct of many years of



treating patients. The time and costs required to develop a new database would exceed the cost of licensing existing data from another provider. In other cases, some companies have gathered sufficient clinical data by performing clinical trials. In these instances, the value of the data can be tied directly to the cost of performing these clinical trials plus a reasonable rate of return.

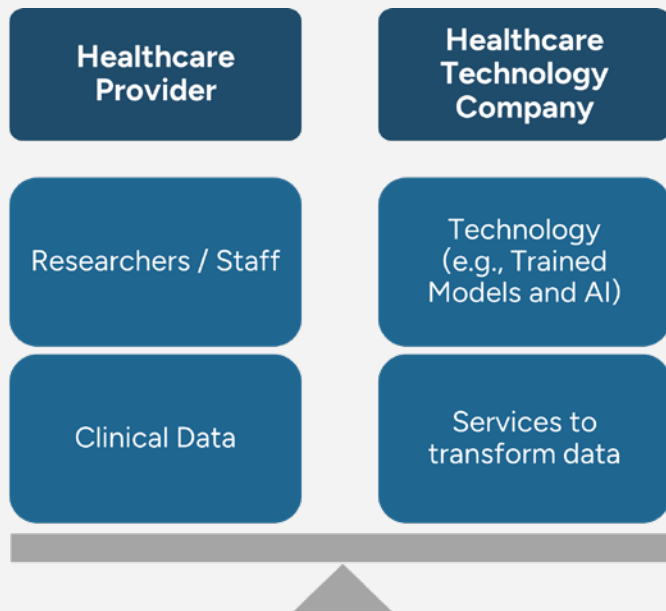
Outsourced clinical trial costs can be used to understand the cost necessary to generate clinical data as both direct costs and market pricing are built into the transactions between study sponsors and clinical trial sites. Clinical research sponsors customarily outsource most of the work of clinical trials to physician practices, hospitals, and clinical research organizations (CROs). Study sponsors compensate clinical trial sites for performance of outsourced services, which can include identifying, enrolling, and obtaining informed consent from a patient subject to actual performance of the intervention or treatment. The publication of the Open Payments datasets over the past 12 years has led to the availability of outsourced clinical research costs for thousands of completed clinical trials for the period from 2013 through 2024. The use of the Cost Approach will ultimately depend on the context and the subject data set being valued.

FMV Considerations

While some data sharing arrangements are simply structured as a payment for use of one party's de-identified clinical data, we have observed other arrangements whereby clinical data may be one component of a larger collaboration agreement between two parties.

An example of this is outlined in **Figure 14**. In this example, one party may provide raw clinical data and the use of its staff and researchers in exchange for receiving a transformed data set that has been de-identified, annotated, and structured. The data licensor may also receive a perpetual license to utilize any trained AI models developed using the data. Because both parties are making contributions, it is important to value each of the contributions that the parties are making and not just the clinical data.

Figure 14: Balance of Collaboration Agreement Contributions



Summary

Companies across the U.S. have been licensing and sharing healthcare data for decades. We believe such transactions will continue to grow in frequency as organizations strive to achieve the quadruple aim of healthcare. With recent advancements in AI, there is tremendous potential for the healthcare industry to improve patient experience, population health and provider experience, all while reducing healthcare costs. Underpinning the development of these tools and insights to achieve these goals is the need for healthcare data.

Companies with data assets should carefully consider the value of their data as they enter into data sharing and data use arrangements or contemplate the contribution of data to joint ventures. Data transactions can be subject to the Stark Law, the Anti-Kickback Statute, and/or private inurement regulations (for non-profit hospitals), where it's important to ensure the compensation paid for healthcare data is consistent with fair market value. Furthermore, most states in the U.S. have also enacted their own fraud and abuse laws that govern self-referrals and fee splitting, which in certain instances may be broader than federal laws and also apply to services covered by private payors. Lastly, the transfer of data assets to a joint venture or de novo entity may have tax implications, making it important to understand the value of the data assets.

Contact Us



David Lo, CFA, ASA

DIRECTOR

dlo@stout.com
+1.303.566.3195

David Lo has been providing clients with analysis and consultation in business valuation, litigation support, intellectual property, and healthcare provider compensation relationships since 2014. His experience includes working with healthcare organizations, attorneys, administrators, providers, consultants, investment bankers, and private equity groups in connection with transactions in the healthcare industry. Many of the transactions he analyzes involve Stark, Anti-Kickback, IRC 501(c)(3), and/or other regulatory implications, which require analysis of fair market value and the determination of commercial reasonableness. Prior to joining Stout, he was a Director with HealthCare Appraisers, Inc.



Bobby Van Dusen, AM

VICE PRESIDENT

bvandusen@stout.com
+1.303.566.3190

Bobby Van Dusen has over three years of experience providing valuation services to both public and private sector clients. He has extensive expertise in the healthcare industry, where his valuation services mainly relate to transaction support, litigation support, and intangible assets. Bobby works closely with attorneys, executives, accountants, investment bankers, consultants, administrators, physicians, and private equity groups in the healthcare space. Many of his analyses involve Stark, Anti-Kickback, IRC 501(c)(3), and/or other regulatory implications, which require analysis of fair market value and the determination of commercial reasonableness. His practice of valuation extends to business valuation, provider compensation valuation, intangible assets, management services arrangements, and marketing arrangements in the healthcare industry. Prior to joining Stout, Bobby was a Senior Associate with Healthcare Appraisers, Inc.



Nick Janiga, ASA

MANAGING DIRECTOR

njaniga@stout.com
+1.303.566.3173

Nick J. Janiga has over 15 years of experience providing his clients with analysis and consultation in business valuation, litigation support, intellectual property, and healthcare provider compensation relationships. His experience includes working with healthcare organizations, attorneys, administrators, providers, developers, consultants, investment bankers, and private equity groups in connection with a broad spectrum of transactions in the healthcare industry. Prior to joining Stout, Nick was a Partner at Healthcare Appraisers, Inc.

About Our Healthcare Capabilities

Stout provides a continuum of valuation, transactional and advisory services to healthcare organizations in order to help them achieve their objectives. As healthcare companies and care providers look to understand the effects of relevant and evolving regulation and the implications of market-driven shifts such as digital health, value-based care, and personalized medicine, we combine deep experience with adept workflows to help organizations navigate the competitive landscape.

Supporting our clients with their most complex issues, our healthcare professionals offer a complete range of investment banking, transaction, accounting and reporting, valuation, and dispute-related services.

Compliance and Risk Management

- Healthcare Valuation -
 - Compensation
 - Business
 - Real Estate
 - Machinery & Equipment
 - Automated FMV Solutions™
- Corporate Integrity Agreement Monitorship
- Gap Assessment
- HIPAA Compliance
- Payment Integrity
- Risk Assessment

Advisory & Consulting

- Physician Compensation Plan Design & Implementation
- Hospital Based Services Optimization
- Value Based Care Plan Design, Oversight & Implementation
- Strategic Advisory

Disputes, Claims, & Litigation

- Ability to Pay Determinations
- Criminal Healthcare Fraud
- Civil Disputes & Investigations
- Damages Opinions
- Expert Witness
- False Claims & Anti-Kickback Cases
- Intellectual Property Claims & Disputes
- Liability Determinations
- Pricing Disputes
- Representations & Warranties Opinions

Transactions

- Buy-Side & Sell-Side Advisory
- Debt/Equity Raise
- Distressed/Special Structures
- Due Diligence (Financial, Operational, Tax & IT)
- ESOP Valuation and Advisory
- Exit Readiness
- Fairness/Solvency Opinions
- Finance Planning & Integration
- Financial Reporting
- Incentive Stock Plans
- IPO Support
- Purchase Price Adjustments
- Purchase Price Allocation
- Quality of Earnings
- Quality of Revenue
- Strategic Advisory



About Stout

Stout is a global advisory firm specializing in corporate finance, accounting and transaction advisory, valuation, financial disputes, claims, and investigations. We serve a range of clients, from public corporations to privately held companies in numerous industries. Our clients and their advisors rely on our premier expertise, deep industry knowledge, and unparalleled responsiveness on complex matters.

Stout Services

INVESTMENT BANKING & RESTRUCTURING

Stout's investment banking team specializes in providing mergers and acquisitions advice, private capital raising, and other financial advisory services.

VALUATION ADVISORY

Stout provides expertise across a broad spectrum of industries in the valuation of business enterprises, complex securities, intellectual property, real estate, and personal property.

DISPUTES, CLAIMS, & INVESTIGATIONS

Stout provides expert testimony and consulting, as well as investigative, claims, and compliance services for financial-related matters.

TRANSACTION ADVISORY

Stout leverages its considerable deal experience to assist clients in making sound business decisions to successfully complete their critical transactions.

ACCOUNTING & REPORTING ADVISORY

Stout delivers a practical, hands-on approach to a wide range of complex accounting topics, such as business combinations, revenue recognition, and more.

SPECIALTY & INDUSTRY SERVICES

Stout offers specialized services focused on helping companies optimize operations, reduce risk, make informed decisions, and enhance value.

COPYRIGHT

© 2025 Stout Risius Ross, LLC

Stout is a trade name for Stout Risius Ross, LLC, Stout Advisors SA, Stout Bluepeak Asia Ltd., Stout GmbH, MB e Associati S.r.l., and Stout Risius Ross Advisors, LLC, a FINRA-registered broker-dealer and SIPC member firm. The terms "Stout" or the "firm" refer to one or more of these legally separate and independent advisory practices. Please see www.stout.com/about to learn more.