Data access, a driver of innovation in the biopharmaceutical sector: a focus on rheumatoid arthritis

By

Tom Achoki

Bachelor of Medicine and Bachelor of Surgery (MBChB)
University of Nairobi, Kenya (2002)

Master of Public Health (MPH)
University of Pretoria, Pretoria, South Africa (2007)

Doctor of Philosophy (PhD)
Utrecht University, Utrecht, Netherlands (2016)

Submitted to the MIT Sloan School of Management
in Partial Fulfillment of the Requirements for the Degree of
Master of Business Administration

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

©2018 Tom Achoki. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly
paper and electronic copies of this thesis document
in whole or in part in any medium now known or hereafter created.

Signature redacted

MIT Sloan School of Management
May 11, 2018

Signature redacted

Anjali Sastry
Senior Lecturer, System Dynamics
Thesis Supervisor

Signature redacted

Johanna Hising DiFabio
Director, MIT Sloan Fellows Program
MIT Sloan School of Management
Data access, a driver of innovation in the biopharmaceutical sector: a focus on rheumatoid arthritis

By

Tom Achoki

Submitted to MIT Sloan School of Management on May 11, 2018
in Partial Fulfillment of the Requirements for the Degree of
Master of Business Administration

Abstract:

The World Health Organization recognizes health data and access to medicines as key ingredients for a well-functioning health system, alongside governance, human resources, infrastructure and financing. Considering the interaction and interdependencies across the various components of the health system, decision makers would need to focus on data trends from different sources within the healthcare system in order to draw instructive lessons to improve health. This is particularly true in the biopharmaceutical industry where safety, efficacy and effectiveness of medicines is a paramount consideration. Collaboration among different stakeholders to improve data access is vital in generating the much needed evidence to drive innovation across the healthcare system. Notably, the biopharmaceutical industry could leverage improved access to data to accelerate discovery of novel medicines, maximize the potential for clinical and commercial success of products and ensure improved access to effective. The shift towards value based care has seen increased demand for data, which has led to the emergence of different models for data access, ranging from data vendors to federations of multiple stakeholders, working jointly to improve the evidence base for decision making. In various settings, federated models, have particularly demonstrated useful attributes that are essential in promoting broader data access within the healthcare ecosystem. However, development of data access models should not be seen as a destination but as a process that is to be continually improved and adapted to meet the needs and demands of a rapidly changing healthcare landscape. Overall, this process should be intimately anchored on the primary objective of improving patient health.

Thesis Supervisor: Anjali Sastry

Title: Senior Lecturer, System Dynamics
Acknowledgements

I would like to acknowledge and express my sincere gratitude to those who assisted in the various steps in the development and finalization of this thesis. Gigi Hirsh and David Fritsche, at Massachusetts Institute of Technology Center for Biomedical Innovation (MITCBI) who gave technical support throughout the process. I also wish to recognize with great appreciation support from Jolly Pradhan, who has been involved in every stage of this process and contributed immensely in troubleshooting various ideas expressed in this thesis. My thesis advisor, Anjali Sastry at the MIT Sloan School of Management, for giving guidance and oversight throughout the entire process. From the MIT Sloan Fellows Class of 2018, Rotimi Adebisiy, Jason Wehrs and Akira Oichi, for their moral support. Finally and most importantly, my sincere gratitude goes to all the key informants, from various organizations who volunteered their time to share their experience and insights to enrich this body of work.
Table of Contents

ACKNOWLEDGEMENTS ..................................................................................................................... 4
INTRODUCTION ...................................................................................................................................... 8
   RATIONALE OF STUDY .................................................................................................................... 12
   SCOPE OF STUDY AND OBJECTIVES ......................................................................................... 13
METHODOLOGY .................................................................................................................................... 15
LITERATURE REVIEW ............................................................................................................................ 18
THE VALUE OF DATA ACCESS: RHEUMATOID ARTHRITIS ................................................................. 21
   WHY DATA ACCESS IN RHEUMATOID ARTHRITIS? ................................................................... 21
   STAKEHOLDERS, THEIR DATA NEEDS AND THE DATA THEY GENERATE .................................. 23
      Patients ....................................................................................................................................... 24
      Health providers ......................................................................................................................... 26
      Payer organizations .................................................................................................................... 27
      Regulatory agencies .................................................................................................................... 28
      Pharmaceutical manufacturers ................................................................................................. 29
   A DATA ACCESS SOLUTION: CORONA REGISTRIES ................................................................... 34
HEALTH DATA ACCESS CONSIDERATIONS ....................................................................................... 38
DATA ACCESS MODELS IN THE BIOPHARMACEUTICAL INDUSTRY .................................................. 42
   DATA VENDORS .............................................................................................................................. 42
   STAND-ALONE ANALYTICS DIVISIONS WITHIN COMPANIES ................................................... 42
   NON-PROFIT LOCAL DATA STORES ............................................................................................. 43
   PRIVATE-VENTURE BACKED DATA AGGREGATORS .................................................................... 44
   FEDERATED MODELS .................................................................................................................... 45
ILLUSTRATIVE EXAMPLES .................................................................................................................. 47
   FDA SENTINEL INITIATIVE ........................................................................................................... 48
   NATIONAL PATIENT-CENTERED CLINICAL RESEARCH NETWORK ........................................... 50
   CENTER FOR HEALTH INFORMATION AND ANALYSIS ............................................................... 52
CONDITIONS FOR SUCCESS AND POTENTIAL RISKS ................................................................... 55
DISCUSSION ......................................................................................................................................... 59
CONCLUSION ....................................................................................................................................... 63
REFERENCE .......................................................................................................................................... 64
APPENDIX A: CLINICAL PICTURE OF RHEUMATOID ARTHRITIS ....................................................... 70
APPENDIX B: GENERAL MANAGEMENT PRINCIPLES FOR RHEUMATOID ARTHRITIS .................. 71
APPENDIX C: PATIENT JOURNEY, BAD OUTCOMES ........................................................................ 73
APPENDIX D: PATIENT JOURNEY, GOOD OUTCOMES ...................................................................... 74
APPENDIX E: COMPARISON OF DATA ACCESS CASE STUDIES .................................................... 75
List of figures:

Figure 1: WHO health system framework ................................................................................................................... 11
Figure 2: Stakeholder mapping, data flows and decisions ............................................................................................. 24
Figure 3: Evidence management throughout the pharmaceutical product cycle .............................................................. 31
Figure 4: Federated data sharing model ....................................................................................................................... 46
Figure 5: Data sharing among different stakeholders in the healthcare system ............................................................... 62
Figure 6: The clinical course of rheumatoid arthritis ........................................................................................................ 70
Figure 7: Prevention and management of rheumatoid arthritis ........................................................................................ 72
Figure 8: Patient journey with poor health outcomes ..................................................................................................... 73
Figure 9: Patient journey with good health outcomes .................................................................................................... 74

List of Tables:

Table 1: Categories of key informants ................................................................................................................................. 16
Table 2: Key questions included in the interview guide .................................................................................................... 17
Table 3: Stakeholder analysis ................................................................................................................................................. 33
Table 4: Comparison of data access models .................................................................................................................... 75
Introduction

According to the World Health Organization (WHO), health information is one of the key six pillars of the health system as shown in Fig.1. The framework outlines the key functions of the health system, as being: health service delivery; development and deployment of human resources for health; collection, analysis and utilization of critical health information; promoting access to medical products, vaccines and other health technologies; ensuring adequate health financing; and effective leadership and governance [1]. Overall, these six components are supposed to interact closely and synergistically to improve population health outcomes.

The WHO framework further elaborates that the intermediate goals of a competent health system are to ensure wide access and population coverage with safe and high quality health services that are necessary to meet the prevailing population health needs [1-3]. It is further emphasizes that this process should be responsive to the legitimate health needs of the population and espouse the democratic principles of equity. Moreover, the health delivery system should be efficient and cost effective, such that persons and households seeking care are not impoverished in the process and that the healthcare system is sustainable in the long-term. Further, a continuous and steady supply of safe and high quality healthcare products (medicines, vaccines and diagnostics) and a robust delivery mechanism (infrastructure, human resources and organization capacity) is essential for the effective performance of the health system [1,2].

Linking all these components together is health information, which is an important pillar of the health system that provides a lens through which to monitor progress and assess performance [1, 3]. Through careful data collection, analysis and reporting, decision makers are able to derive useful insights on how various parts of the health system are interacting with each other and identify critical pain points in the healthcare delivery pathways in order to institute corrective action [1, 3, 4]. The systematic assessment of health system performance is fundamental in efforts to create and deliver value to the patient and society in general.

For instance, careful examination of data on patient health outcomes covering mortality and morbidity trends could allow health system stewards to determine priority health needs in a given population so as to objectively allocate resources [5, 6]. Further, combining the information on health outcomes with individual utilization levels of essential medicines and other health services, would reveal critical access gaps in delivery and direct policy attention accordingly. This learning
cycle is further extended, in cases whereby data on health outcomes, resources (financial, human and material) and levels of health service coverage are combined, allowing decision makers the opportunity to distill additional insights into the efficiency and cost effectiveness of healthcare delivery [1, 6, 7].

Invariably, continuous monitoring and evaluation of data from the various sections of the health system, empowers decision makers to appreciate and anticipate the population health needs and respond accordingly. Discerning decision makers operating in such adaptive and learning health systems, are able to draw instructive lessons which would allow them to individually or collectively optimize the impact of their actions in health service delivery value chain [2, 6-8].

On the other hand, there are many pitfalls when decision makers pay attention to data from only one component of the health system without accounting for trends and interactions with the other related parts of the system [8, 9]. It has been demonstrated that this narrow approach often results in spurious conclusions. For instance, studies from select health systems have shown that investment in health system inputs and processes does not necessarily translate to expected outcomes if contextual factors are not considered and addressed [10, 11]. Therefore, the utility of data in health system performance assessment is maximized when it is assessed collectively, rather than in isolated siloes.

In addition, to avoid potential blind spots, it is fundamentally vital for decision makers to carefully examine trends across the different components and levels of the health system such that a consistent and instructive story emerges. For example, if decision-makers only focus on the national aggregate estimates of performance, they might miss out on important subnational trends that could reveal geographical inequalities, requiring closer policy attention [12]. This is clearly illustrated by Olives [13], where high heterogeneity in the distribution of hypertension at both state and county levels in the US was recognized.

These examples demonstrate that in order to make sustainable progress, decision makers must seek to have a holistic understanding of the various interactions across the health system that determine patient outcomes, rather than focusing on piecemeal information generated from isolated data siloes. Broader insights derived from triangulation of multiple data sources from different stakeholders within the health system provide unique opportunities for continuous innovation to improve population health.
Unfortunately, in many healthcare systems in different parts of the world, the processes linked to data collection, analysis and reporting, are highly fragmented [14, 15]. This fragmentation effectively curtails the ability to generate nuanced system insights required to catalyze and drive innovation to create value for patients and society at large.

In this thesis, we take the perspective of the biopharmaceutical industry as a stakeholder operating within the broader health system. This perspective is anchored on the WHO framework that stipulates that a well-functioning health system ensures broad access and rational use of quality, safe, efficacious and cost-effective medical products, vaccines and other health technologies to improve population health [1, 2]. This effectively means that the biopharmaceutical industry has a clear responsibility and mandate to discover and develop innovative medical products and other health technologies that improve the health of patients.

In our analysis, we seek to determine how improved data access would catalyze and drive innovation with the ultimate goal of creating and delivering value to the patient and society at large. Value within the health system context, is defined as the level of health outcomes achieved accounting for the resources spent [16]. Meanwhile, innovation is broadly defined as the application of new or altered products, services, processes, systems, policies, organizational structures, or business models that create value for the patient. Therefore, within the context of biopharmaceutical industry, innovation can be taken to mean the discovery and development of effective pharmaceutical products which prevent, cure or improve a disease condition; as well as reduce mortality or morbidity. In addition, innovation encompasses those interventions that reduce the cost of care; improve the quality of life; are safer or easier to use; or improve patient compliance and persistence [15-17].

This thesis are organized as follows; in the ensuing subsection, we describe the rationale and objectives of the study. In section 2, we cover the research methodology employed in this paper. In section 3, we present a brief literature review highlighting data access as a basis for value creation within the healthcare system. Section 4, presents a motivation for data access within the rheumatoid arthritis therapeutic area, with Corrona registries, as an illustrative example. In section 5, we cover the health data access considerations based on insights gained from interviewing the key informants. Section 6, presents data access models within the biopharmaceutical industry and with specific illustrative examples following in section 7. In section 8, we highlight the conditions
necessary for success and potential risks to data sharing. Finally, in section 9, we discuss some of the key considerations decisions makers need to be aware of and conclude with recommendations.

Figure 1: WHO health system framework

Rationale of study

The biopharmaceutical industry is a key stakeholder in the healthcare delivery value chain charged with the responsibility of developing and delivering safe and quality medicinal products essential for health provision. In the WHO framework [1-3], effective population coverage with medicines, vaccines and other health technologies is recognized as one of the six pillars of a well-functioning health system [1, 2]. Health data on the other hand, provides the basis for assessing the performance of the various components of the health system.

Therefore, in order for health system decision makers, to be fully aware and effectively respond to the population health needs, data access to recognize gaps, opportunities and generate evidence to guide action is fundamental [1-3]. For example, decision makers in the biopharmaceutical industry would need to leverage data in the discovery, development, marketing and distribution of safe, efficacious and effective pharmaceutical products. Further, the identification of unique biomarkers, new drug targets, and characterization of patient heterogeneity across subpopulations are some of the useful insights that could be gleaned from various data sources and benefit clinical development of medicines.

Regulatory and access considerations for safety, efficacy and cost-effectiveness could be satisfied by innovatively leveraging health data from different sources within the healthcare system. This is particularly relevant given the recent trends in the healthcare landscape where the focus is more on delivering value to the patient with measurable outcomes, rather than the volume and quantity of services provided [16, 17]. The focus on quality of care outcomes and other aspects of comparative assessment, also exerts more data demands to the health system stakeholders, which cannot be satisfied by a single entity or stakeholder [17, 18, 19].

This is all happening in the context whereby, in many parts of the world, the patient is becoming more informed, discerning and empowered to take an active role in care decision making rather than being a passive consumer [20, 21]. Payers and regulatory organizations are increasingly looking for demonstrated effectiveness of medicines from real world settings, in addition to clinical trials, in order to make authorization and reimbursement decisions [21-23]. This is particularly true in the face of a steady pipeline of highly-priced specialty drugs that are potentially curative for chronic conditions. This trend is coupled with downward cost pressures on prices to
ensure affordability in the face of dwindling budgets; and limited understanding of the effectiveness of medicines in real world situations to justify coverage and reimbursement decisions. All these factors are instrumental in driving the trends towards better evidence generation with the health system landscape [17, 22, 24].

The concept of value based healthcare delivery espouses the fundamentals of evidence based medicine that places the patient at the center. In this framework, all stakeholders work in tandem to ensure that all actions are geared towards improvement of health outcomes. However, given the complex interactions that characterize the patient journey across the healthcare system, solutions to some of the challenges that hinder the attainment of optimal patients outcomes cannot be in the purview of a single stakeholder, rather than the collective [1, 2]. In this dynamic landscape, access to data provides the common currency that facilitates meaningful communication and interaction across the health system [1-3].

This is achievable, by leveraging the recent technological developments, including the capacity to store and process vast amounts of data from multiple sources. It further presents a unique opportunity to facilitate system wide learning aimed at improving the service quality and efficiency of healthcare systems. In addition, actors within the biopharmaceutical industry that are keen to accelerate development and create value for their patients and society at large, stand to benefit immensely from these developments [17, 23, 25].

Scope of study and objectives

Broadly, this study seeks to answer a set of fundamental questions that revolve around data access among different stakeholders involved in the healthcare delivery pathways for chronic conditions, taking the example of rheumatoid arthritis. In this study, we seek to understand some of the priority questions that decision makers from among different stakeholder groups within the health system have in their quest to improve health. We further describe different models for data access, covering some of drivers and barriers to improved access.

More specifically the study will have the following objectives:

- To describe data access models within the healthcare landscape
- To identify some of the priority questions and decisions by different stakeholders
involved in the health delivery value chain

- To highlight some of the data access considerations in the quest to improve patient outcomes and drive innovation

- To recommend measures to improve data access among health system stakeholders
Methodology

This study entailed a literature review covering, a subset of published articles and gray literature (including working papers and policy documents) on health data access within the healthcare landscape. In as much as we considered data from different health system stakeholders, our search specifically focused on those publications that were relevant to the biopharmaceutical industry.

Briefly, our search strategy consisted of Google, the PubMed and Google Scholar, and covered a subset of articles from January 2008 to January 2018, limiting to English-language publications and omitting the publication types of “letters,” “patents”, and “comments.” The year 2008 was arbitrarily selected to serve as a baseline to ensure extraction of insights that are reflective of the current trends in the healthcare industry. We also felt that a ten year timespan is sufficient to explore the latest temporal trends.

Search terms included “data access in biopharmaceutical industry”, “data access in rheumatoid arthritis”, “data collaboration models in biopharmaceutical industry” and “data sharing in rheumatoid arthritis”. The abstracts and summaries of the identified publications were assessed for relevance to the topic and a subset of those deemed satisfactory were selected for a more detailed review.

We excluded articles and reports that described data generated in the preclinical stages of biopharmaceutical research (since this was considered out of scope in our study). Our review primarily focused on articles that reported on data generated throughout the product life cycle, starting from clinical trials to the real world utilization. This included, but not limited to data generated for regulatory approval and post authorization surveillance purposes, health insurance coverage and payment options, service utilization and patient experience and outcomes.

The initial review provided, practical insights into the treatment pathway for rheumatoid arthritis patients, and their broader interaction with the different healthcare system stakeholders. Further, we were able to obtain information on the key questions and decisions by the key stakeholders involved in the healthcare delivery value chain for rheumatoid arthritis patients. Through this process, we were able to build a comprehensive picture of who are the main stakeholders within the value chain as well as their influence on the broader health system in terms of value creation for the patients.

The second phase comprised of a description of various methods through which biopharmaceutical
companies access data. Through expert opinion and literature review, we selected different case studies that were relevant to our research and described them in detail. Cognizant of the fact that data access models vary in degrees and levels of access, we purposely sought to select case studies that are representative of the wide spectrum.

In the third step of the research, 15 key opinion leaders (KOL) identified (in the first and second stages) as public experts that had published articles on the topic of data access, were conveniently selected as key informants to give further perspective to a set of open ended questions related to the topic. Table 1, highlights the different categories of key informants that were consulted. These comprised of experts in the biopharmaceutical industry, respected management consulting practices, research and academia, health insurance and provider organizations. A key informant guide was used to give structure to the interview which was recorded for subsequent transcription.

Table 1: Categories of key informants

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biopharmaceutical industry</td>
<td>5</td>
</tr>
<tr>
<td>Management consulting</td>
<td>3</td>
</tr>
<tr>
<td>Research and academia</td>
<td>3</td>
</tr>
<tr>
<td>Health insurance</td>
<td>2</td>
</tr>
<tr>
<td>Health provider organizations</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>15</strong></td>
</tr>
</tbody>
</table>

The interviews were conducted between 3rd April 2018 and 17th April 2018. Briefly, arrangements were made to secure a 30-min appointment and a suitable venue to conduct the telephone or face to face interview. This ensured that participants were not unduly distracted during the interview. Before the interview started, the researcher introduced himself, explaining the objectives of the study and securing verbal informed consent to proceed with the interview. Participants were made aware that they could cease participating in the interview at any stage without prejudice. Leading questions, prepared by the authors, ensured that the participants responded to the key topical issues of interest. Table 2, shows some of the key questions that were included in the interview guide.
Table 2: Key questions included in the interview guide

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td>What are the main factors that determine access to data in the healthcare landscape?</td>
</tr>
<tr>
<td>b)</td>
<td>Are you aware of examples/models through which the actors within the biopharmaceutical industry access data from the broader healthcare system?</td>
</tr>
<tr>
<td>c)</td>
<td>What are the key factors that facilitate data access within the healthcare system?</td>
</tr>
<tr>
<td>d)</td>
<td>In your opinion what are the factors that hinder broader data access among different health system stakeholders?</td>
</tr>
<tr>
<td>e)</td>
<td>What steps could decision makers within the health system make to bridge the prevailing data access gaps?</td>
</tr>
<tr>
<td>f)</td>
<td>What value could be created by the biopharmaceutical industry through improved data access?</td>
</tr>
</tbody>
</table>

Interviews were transcribed and analyzed applying thematic content analysis. This approach is well-suited for semi-structured interviews as it is used for coding text with a predefined coding system which can then be refined and completed with new themes emerging. Some of the coding categories that were identified a priori from the literature on data access, include; “data demand”, “data supply”, “data access barriers” and “success factors”.

Data were both deductively and inductively coded, whereby a series of codes were developed and then grouped into similar concepts. These concepts were then combined to form categories as the main determinants of data access. Further from the interviews, additional insights were derived on the classification of data access models within the biopharmaceutical industry, which were validated from the literature. Finally, additional categories which included, “determinants of success” and “barriers to progress” emerged as recurrent considerations to ensure success of any data access models.

After data collection, emerging themes were identified and categorized for manual analysis. There was further consultation with the KOL to obtain additional insights and validation of the findings from the previous stages. Finally, the information collected from the previous stages was synthesized into a thesis with policy recommendations.
Literature Review

Evidence generation and management throughout the various stages, from pre-clinical research, development and commercialization of a pharmaceutical product is vital to ensure that patients have access to the much needed therapeutics to improve health. The United States Government in its 21st Century Cures Act (Cures) [26, 27], recognizes the important role of data access in the discovery, development, and delivery of new treatments in order to maintain America’s global status as the leader in biomedical innovation. Some of the steps proposed in Cures include, modernizing clinical trials and the means by which safety and efficacy data is accumulated and analyzed. Other measures include broader and more collaborative development, leveraging disparate data sources among stakeholders, and utilization of biomarkers, which help assess how a therapy is working for different subpopulations earlier on in the development process [27, 28].

Cures, further espouses a patient centric care model and aims to help improve delivery by ensuring that electronic health record systems are interoperable for seamless patient care and help fully realize the benefits of a learning health care system [26-28]. The Pharmaceutical Research and Manufacturers of America (PhRMA) [29], in support of the proposals laid out in Cures, recommends the use of real-world evidence and observational data to establish the benefit of medicines. This approach would be particularly relevant for objective identification of new indications for a product that has already demonstrated safety and efficacy for another use through standard clinical trials.

PhRMA, further advocates for enhanced collaboration between government agencies, the biopharmaceutical industry, and academia to transform the clinical trial ecosystem, as way to accelerate development and improve access [29]. In building the case for stakeholder collaboration, WHO [30] and Institute of Medicine [31], also assert that data sharing can accelerate new discoveries by avoiding duplicative trials, stimulating new ideas for research, and enabling the maximal scientific knowledge and benefits to be gained from the efforts of clinical trial participants and investigators.
However, Cognizant [32], cautions that randomized clinical trials, the established way for clinical development of therapeutics, are increasingly being challenged, with key stakeholders such as payers and regulators requesting additional evidence to validate the claims that pharmaceutical products are beneficial in the real world settings. These requests for value demonstration in the real world setting can only be derived by leveraging observational data obtained from other sources such as routine clinical practice and patient registries.

In support, the Network for Excellence in Health Innovation (NEHI) [17], suggests that evidence derived from real world data sources is a potentially transformative force in the US healthcare system. By merging data from multiple sources to generate real world evidence, NEHI, postulates that this would unleash a new era of healthcare innovation and improve patient outcomes. It is further suggested that by tapping into different data sources, such as clinical, genomic and socioeconomic trends, useful insights into the patient’s characteristics would be identified that would help guide and tailor appropriate treatment options.

Deloitte [25] and NEHI [17], further identify the utility of observational data at the various stages of the product development and commercialization. For instance, identification of unique patient characteristics would help expedite the hypothesis generation and bring a clear focus in the clinical research and development processes. This could include the design of clinical trials and recruitment of patients that optimize the process. In addition, combining real world data from various sources would be essential in making the case for indicator expansion, reimbursement justification as well as better insights into the safety of the products in the long-term. This point is further supported by, McKinsey & Company [21], which suggests that researchers can mine the data to see what types of treatments are most effective for particular conditions, identify adverse effects, and gain insights that can help improve health outcomes and reduce healthcare costs.

Considering the shift that healthcare systems across different parts of the world are making towards value based care, Hisey and Davis [33] advocate for end to end management of evidence, using data from multiple sources as a way to accelerate innovation and improve patient outcomes. The authors further suggest that this is particularly true in the emerging healthcare landscape, where development of effective treatments is more personalized requiring a deeper understanding of the
patient's characteristics, including biomarkers, comorbidities and treatment pathways. Therefore, for successful development of effective therapeutics, a greater amount of evidence from multiple sources is vital, making collaborative data sharing a paramount consideration.

Overall, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and International Society for Pharmacoepidemiology (ISPE) [34, 35], recognize the utility of real world data from various sources as useful in the generation of evidence for decision making in the healthcare space. However, they provide guidance that, for the evidence to be useful and acceptable by different stakeholders, it must be reproducible, transparent, including data sharing so that it can be reused, as well as ensure methodological soundness and inclusivity of different stakeholders.

From this brief literature review, it is clear that various authors and organizations recognize that improved data sharing among different stakeholders would be help accelerate innovation within the biopharmaceutical industry and ultimately create value for the patient. In the next section, we describe the value of data access in the rheumatoid arthritis landscape, and provide an illustrative example.
The value of data access: Rheumatoid arthritis

It is clear that data access has become a central theme in the on-going debates on evidence based practice that is necessary to ensure effective patient centric and value based care models. As a result a variety of different models for data access within the healthcare ecosystem have emerged to address specific information needs for different stakeholders [21, 31]. Considering that data access in this case is primarily useful to provide evidence to guide improvements in patient health, focusing on a specific therapeutic area or disease condition is crucial, in order to highlight the key specifics related to data access.

In this section, we focus on rheumatoid arthritis and map out the various stakeholders in the healthcare ecosystem, highlighting their data needs, decisions and actions that they have to take in order to improve health. After the stakeholder analysis subsection, we further describe a data access mechanism, Corrona registries, that has emerged in this specific therapeutic area to respond to the stakeholder’s data needs. We briefly highlight some of the innovation and value creation potential derived from this data access model.

Why data access in rheumatoid arthritis?

Rheumatoid Arthritis (RA) is chronic condition that can result in joint deformities and disability and ultimately a reduced life expectancy. According to a recent study, the population suffering from rheumatoid arthritis around the world was estimated at 21.3 (95% CI: 19.5-23.6) million people in the year 2016 [37].

In as much as the disease primarily involves the joints, it also has extensive extra-articular manifestations, including vital organs such as the circulatory and pulmonary systems where it could result in significant and serious comorbidities. There are various factors that are important in determining the onset and prognosis of the disease, including genetic factors and lifestyle choices such as smoking and diet [38]. The clinical progression of the disease and its determinants is covered in Appendix A.

If not well controlled, rheumatoid arthritis often leads to joint destruction, disability and compromised quality of life. Long term complications include, cardiac and pulmonary disease as well as increased predisposition to infections and malignancies [38, 39]. Effective management as shown on Appendix B entails, primary and secondary prevention measures such as smoking
cessation and weight management, as well as early treatment with disease modifying anti-rheumatic drugs.

Management measures for rheumatoid arthritis can only be effective when there is broad data access among different stakeholders throughout the continuum of care. This is because effective care is often integrated and multidisciplinary and should cover different aspects including, management of patient’s risk factors (genetic, socio-economic and lifestyle); diagnostic and treatment options (including severity, comorbidities and adherence) as well as cost and health outcomes considerations [38-40].

As previously mentioned, an in-depth understanding of genetic variability across different subpopulations could be instructive in guiding discovery of effective treatments that are best suited for certain population groups. Further, by pooling the data from various patients across different treatment centers, registries and geographies, optimal patterns of care could be identified and offer a powerful basis for development and testing novel treatment products and algorithms that would ultimately improve patient outcomes [17, 24]. In addition, given the chronic course of the disease, rheumatoid arthritis patients are often faced with multifaceted challenges [38, 39], which could range from financing constraints to low treatment adherence; factors that could affect treatment outcomes. A good understanding of all these important factors that may affect treatment outcomes for rheumatoid arthritis patients can only be achieved through proactive measures to share data among different stakeholders [40].

For illustration, in Appendix C, shows a single patient journey with bad health outcomes, highlighting some of the gaps in data sharing that were responsible for the poor outcomes. In contrast, Appendix D shows a patient journey with good health outcomes, and illustrates how data sharing across the care pathway has optimized the outcomes. Overall, aggregating the experiences of individual patients across the continuum of care could provide significant lessons for a learning health system that is keen to improve health outcomes for rheumatoid arthritis patients. Improved data sharing among healthcare system stakeholders allows for the generation of useful insights leveraging multiple sources, and thus enabling the healthcare system to become more preventive and predictive through better patient risk stratification as well as more effective by targeting interventions appropriately [17, 25, 31].
Stakeholders, their data needs and the data they generate

In order to fully appreciate the various data sources that are relevant in the healthcare ecosystem, it is vital to map out the key stakeholders. Normally, when seeking healthcare, the patient interacts with various healthcare providers, who use medicines and other health technologies that have been manufactured by biopharmaceutical companies with the approval of regulatory agencies [17, 21, 25]. Payment for health services is the responsibility of payer organizations or in some cases out of pocket from the patient or a combination of both. Therefore, in our case we recognize the key stakeholders as, patients, health providers, payers, regulatory agencies and biopharmaceutical manufacturers [20, 22, 24]. In this section, we describe the different health system stakeholders, highlighting some of their priority questions, decisions and actions that they have to make in order to ensure optimal health outcomes for those seeking care. Figure 2, is a schematic presentation of the interaction of the different stakeholders.
Patients

Ultimately, when a rheumatoid arthritis patient is seeking healthcare, their key questions mainly revolve around whether the care they are receiving will alleviate their symptoms, restore their function and improve their health and general well-being. They would also want to know what specific actions they would need to take in order to realize optimal health outcomes. There is often
uncertainty in choosing treatment regimens given the variable severity and course of the disease among different patients [38-40].

Recent technological advancements have greatly enhanced the capacity of healthcare systems around the world to anticipate and respond to the specific needs of individual patients [21, 25]. In patient-centric models, care is more personalized and preventive in nature. It espouses partnership and collaboration among various stakeholders working together to optimize care. At the center is the patient, empowered to take a proactive role in making crucial decisions related to the care they are receiving.

Therefore, full awareness of the patient unique characteristics such as genetics, diet, life style choices, and socio-economic factors, are vital in the effort to tailor personalized interventions that would yield the desired outcomes. By routinely collecting and analyzing patient data, decision makers are able to take appropriate actions, effectively monitor progress and intervene in a timely manner whenever necessary. For instance, tracking of the patient data on adherence to treatment could be useful to guide decision makers to explore the various factors that are responsible such as adverse reaction or financial constraints so as to intervene appropriately.

Further, when appropriately pooled together, data derived from various sources including, electronic medical records, claims and billing data, product utilization data and adverse events as well as disease registries, are all essential in providing real world evidence to optimize care for the patient [25, 27, 28]. Ranging from expedited discovery of suitable treatment options for specific subgroups to identification of public health measures to ensure access to medicines, improved access to patient data is a vital precondition for progress.

In addition, patient-generated data including, behavioral, lifestyle and occupational trends, when used in combination with other sources, could be useful in providing a fuller picture into the patient’s risk factors and health status [21, 25]. For instance, data collected from mobile devices could prompt a healthcare provider to question why a rheumatoid arthritis patient is having reduced mobility or reduced adherence to medication. This could be due to a myriad of factors such as a flare-up of the disease, depression or injury- all requiring different interventions. Retail data on purchases, diet and lifestyle choices could also point to important risk factors such as smoking,
alcohol abuse and obesity, all that have been known to accelerate the progression of rheumatoid arthritis.

Social media is also emerging as a valuable source of data to identify patient sentiments, as it relates to the care they are receiving and the products they are using [20-23]. Technological advancements in data curation and analytics is making it possible to gather and interpret this data in meaningful ways. Another source of data on patients is coming from patient advocacy groups. Partnership with such organizations, would yield valuable data to guide care [21, 25, 27]. Invariably, there is compelling evidence which proves that listening to the patient’s voice can result in real benefits to the patient’s survival and general well-being.

**Health providers**

Normally, the primary consideration of a health provider is to select the best treatment options, based on the best available information to optimize the patient’s outcomes. Providers have the first hand interaction with the patient throughout the care seeking journey and make important decisions in influencing the quality and quantity of information available [15, 17, 21]. By definition, the healthcare provider category encompasses persons and businesses that provider health services to patients. This includes practitioners such as nurses, dentists and doctors as well as the organizations within which they work, such as clinics, hospitals and diagnostics centers [1-3].

Within the course of routine care, healthcare providers often collect important data about the patient’s condition, which includes but not limited to, signs and symptoms, family and social history as well as physical examination findings. They also order various diagnostic investigations and prescribe treatments, and monitor outcomes both in the short and long term. All this information is often stored in a patient file which could be available as an electronic medical record [21, 23, 25]. For some specific conditions, this could also be available through specialized registries, where specific patients are longitudinally tracked. Therefore, health providers play a critical role as a source of data collection within the healthcare system.

When carefully assessed in combination with other sources, data collected by providers could be useful in determining the effectiveness of different treatment regimens for specific conditions and patient groups. This is particularly relevant to a condition such as rheumatoid arthritis which often
has variable severity and prognosis among different patients groups [38-40]. Such data could also be useful in identifying patterns related to adverse events related to specific treatments, requiring prompt interventions, and hospital readmissions which could instructive to the care provider teams that are keen to optimize care.

For example, combining insights into treatment effectiveness with the cost of various treatment regimens could provide decision makers perspective into cost effectiveness and efficiency of the different care options [1, 2, 25, 33]. This demonstrates the imperative for seamless data pooling from different stakeholders in order to generate robust evidence to advance public health.

With the movement towards value based care delivery, where the focus is on patient outcomes rather than quantity of services delivered, many providers have seen the deliberate expansion of rigorous electronic medical records as a tool to collect data points that capture the most important metrics for patient care. Ultimately, this wealth of information if properly harnessed could empower healthcare providers to make better treatment decisions and monitor patients effectively [17, 21, 25]. In addition, other stakeholders, such as biopharmaceutical companies could have timely and real world insights into the performance of their products and catalyze useful innovations by leveraging provider generated data.

**Payer organizations**

Ultimately, payers are keen to reimburse treatments and care options that improve patient outcomes at a lower cost [21-23]. Value based care, that focuses primarily on outcomes is emerging as a dominant trend in the health insurance industry. In addition to the efficacy of medical products demonstrated in controlled clinical trial settings, payers are increasingly wanting to see demonstrated clinical impact in the real world setting, in order to make reimbursement decisions [15, 17, 33]. Therefore, there is increasing pressure on other stakeholders such as biopharmaceutical companies and health providers to demonstrate value for their products or services in order to justify reimbursement.

In order to meet these value based requirements, stakeholders need to have access to data on treatment effectiveness from multiple parties, including clinicians, regulators, and researchers [17, 25, 33]. In this quest, there is a compelling need to combine information on effectiveness with
costs of treatment in order to make a meaningful comparative assessment in terms of costs and effectiveness of different treatments.

Given the amount and types of data in their custody, payers have the capacity to play an important and leading role in advancing the evidence base for value based care [21, 24, 25]. Payers collect data on type of treatment, patient response and cost, all which are crucial pieces of data to consider when evaluating which treatment regimen to adopt for a specific patient. Claims data can clearly map out the patient’s healthcare seeking patterns which is important for chronic conditions such as rheumatoid arthritis. For example, assessing data on the amount of health insurance copayments, treatment adherence and health outcomes, could point to a patient who is defaulting treatment as a result of financial pressure and hence the need for appropriate intervention.

Certainly, innovations such as outcome or performance based risk sharing schemes that are emerging in various therapeutic areas in the face of high costs of novel therapeutic agents are only feasible in a data rich environment [17, 21, 23]. In this case the performance of a product is tracked in a defined patient population over a specified period of time and the level of reimbursement is determined by the achievement of some specific end points that have been agreed upon (normally by the payer and the biopharmaceutical company).

Overall, payers have an important role due to the amount of data and information generation capacity at their disposal. Therefore, if payers could readily leverage their data with other stakeholders in the healthcare ecosystem, they have a strong opportunity to improve the depth and quality of available evidence to make decisions and drive innovations that improve patient outcomes.

**Regulatory agencies**

The primary role of regulatory agencies is to steward and guide the process of approving safe, quality and efficacious health products for use in the population that would derive evidence based health benefits [17, 31-33]. Normally, this process is accomplished through careful monitoring of data that is generated through various stages of clinical trials by the product sponsors. To fulfill their statutory responsibilities, regulatory agencies have stringent guidelines that carefully balance the risks and benefits to the patients (or trial subjects) and guarantee safety of the population.
Increasingly, there has been a push to expand the data used to gain regulatory approval to include real world data that is generated from routine care and other observational studies related to the product [17, 21, 25]. This is particularly true when considering that in clinical trial settings, patients and study specific conditions are carefully selected to optimize the observed effect of the product, which is not nearly reflective of the real world conditions where the patients use the products once approved [25, 31].

Pharmacovigilance and post-marketing safety surveillance are also other data sources that are vital in the regulatory process [21, 25, 32]. Regulatory risk management procedures for medicines and other healthcare products may include a requirement to monitor previously documented or undocumented adverse events to gain further insights. Periodic screening of product users to avoid complications, or the mandatory enrollment in a registry for closer monitoring are all risk management measures linked to the regulatory process [18, 20]. Additionally, full-fledged epidemiological studies maybe required if safety signals are identified from a product that is already in the market, e.g. from communication and observations of healthcare providers.

Creating an enabling environment where all these data sources that are generated throughout the regulatory process could be seamlessly shared among stakeholders, is vital in driving innovation in the industry and creating value for the patient. For example, a positive step towards this is value creation is when the Food and Drug Administration (FDA) introduced a requirement for clinical trial sponsors of regulated products to register studies and report results at ClinicalTrials.gov [31].

**Pharmaceutical manufacturers**

In general, biopharmaceutical companies are mainly concerned with the development of health products that have demonstrable benefits (in comparison to risks) to patients [21, 25]. Companies are also keen to ensure that the products that they produce yield a fair return on investment. Manufacturers have a particularly special role in the process of medicine discovery and production, because they are ultimately responsible for the data that gets the drug approved [17, 18, 25]. In addition, with the expectation that responsive industry actors would give support “beyond the pill”, companies are keen to understand patient and prescriber behavior and preferences that ultimately determines the utilization and health outcomes [20, 22, 23].
In this context, biopharmaceutical companies (particularly those with products attracting high price points) have been hard pressed to demonstrate that their products work in real-world conditions prior to reimbursement [17, 18, 21, 25]. Through careful analysis, data from various real-world sources can play an important role in proving the value of a new drug and reimbursement justification. Long-term efficacy and safety studies conducted under adaptable and pragmatic real-world conditions, could provide a basis through which biopharmaceutical companies effectively negotiate performance based reimbursement schemes with payers-accelerate market access and other commercialization measures. In addition, adoption of innovative strategies that integrate patient and investigator level data can ease costly regulatory pressures, allowing cash-strapped companies to extend those resources and generate more robust trial initiation plans [25, 33].

The clinical development process has to clearly demonstrate safety, efficacy and effectiveness of the medicinal product prior to obtaining approval. Normally, the patient populations involved in the clinical trial process, gradually increases as the candidate product passes various rigorous stages of clinical development [31, 33]. This process is lengthy and costly, and has lately seen calls for adaptations that would accelerate clinical development and access to medicines while ensuring safety and effectiveness. Improved data access among different stakeholders across the health system could yield tremendous benefits in terms of optimization of the clinical development process [21, 25].

For instance, by clearly understanding the patient characteristics and determining who is the most likely to benefit from the pharmaceutical product, companies are better able to design appropriate trials and select suitable participants which would allow for a speedy demonstration of efficacy and effectiveness. Further, when clinical teams incorporate data from electronic health records, patient registries, prescription records, and other real-world sources, they can achieve great benefits, such as making better protocol designs and speeding patient recruitment for clinical trials [21, 25, 32]. This could lead to a faster navigation of the regulatory process and lowering the costs associated with clinical trials particularly through adaptive processes that leverage Bayesian statistical models [21].
Figure 3, summarizes the various stages of the product life cycle where data sharing among stakeholders could lead to optimal results. Briefly, earlier on in the discovery stages, pooling of data from multiple sources could help accelerate the identification of appropriate biomarkers and promising molecules to advance to the next stages without incurring excessive costs.

Real world evidence derived from a combination of multiple data sources could also be effectively leveraged to support label expansion to include other indications where the product is beneficial [21, 25]. In this case, real world evidence enables products sponsors to build a case for authorization and potentially avoid having to conduct additional clinical research as part of the approval process.

Finally, understanding the demand and supply factors associated with a product is essential in ensuring a successful access and commercialization strategy in any given market. Invariably, broad access to data sources within the healthcare system, would be useful in providing insights into various market dynamics for successful access and commercialization of products [20, 22, 23].
Furthermore, public health considerations such as burden of disease, procurement and supply chain factors, prescriber patterns and preferences, patient engagement and adherence among others are all useful factors to consider, for industry actors that are keen to improve patient health.
Table 3, below summarizes the key questions and decisions that various stakeholders have to make, highlighting some of the data available.

Table 3: Stakeholder analysis

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>Some priority questions</th>
<th>Decisions and actions</th>
<th>Data available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>What interventions or measures do I need to take to improve my health?</td>
<td>Seek and adhere to appropriate treatment and lifestyle choices</td>
<td>Consumer and lifestyle data e.g. diet, alcohol consumption, adherence to treatment etc.</td>
</tr>
<tr>
<td></td>
<td>Am I getting the best care possible to improve our health condition?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biopharmaceutical</td>
<td>Are our products safe, efficacious and effective?</td>
<td>Develop and test new treatments through rigorous scientific studies</td>
<td>Preclinical and clinical development data for various products</td>
</tr>
<tr>
<td>companies</td>
<td>How do we demonstrate value for our products in this competitive landscape?</td>
<td>Support patient access to treatment</td>
<td>Specific disease/product registries, pharmacovigilance data etc.</td>
</tr>
<tr>
<td></td>
<td>How do we differentiate our products in a competitive market?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Payers</td>
<td>Are the products we are reimbursing cost effective?</td>
<td>Reimbursement decisions for treatments with demonstrated value</td>
<td>Coverage and reimbursement processes; claims data for various indications e.g. costs</td>
</tr>
<tr>
<td></td>
<td>How can we contain the escalating costs associated with healthcare provision?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Providers</td>
<td>What do we do to improve the health of our patients?</td>
<td>Make appropriate treatment and prescription decisions based on the best available evidence</td>
<td>Health utilization data, patient outcomes data, Healthcare expenses etc.</td>
</tr>
<tr>
<td></td>
<td>Why are our patients having different treatment outcomes despite our best efforts?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulators</td>
<td>What evidence do we need to approve safe and effective treatments?</td>
<td>Approve the treatments that have been demonstrated to benefit patients</td>
<td>Regulators normally request data from product sponsors (biopharmaceutical companies) throughout various stages of the approval process</td>
</tr>
<tr>
<td></td>
<td>What do we need to do to ensure that patients have faster access and benefit from innovations in life sciences?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A data access solution: Corrona registries

Corrona operates the largest real world observational database for rheumatoid arthritis and includes registries for other autoimmune diseases, such as psoriasis, psoriatic arthritis and spondyloarthritis, multiple sclerosis and irritable bowel disease. Its rheumatoid arthritis registry sites are expansive and cover most parts of the United States, with inclusion of additional sites in Japan and other countries in the pipeline for subsequent inclusion. This allows the firm to have a strong position operating as a real world evidence business that provides registry data and consulting services to biopharmaceutical companies. However, despite having a close working relationship with industry, Corrona remains an independent registry without any ownership links to the biopharmaceutical industry [36, 41].

Key stakeholders

Corrona was founded in 2001 by leading rheumatologists keen to improve care of patients with autoimmune diseases. Participating clinical sites, which include private and academic medical institutions with the capacity to conduct research contribute data from their routine clinical encounters of patients with specific conditions to the Corrona registries. Participating sites are compensated for submitting completed questionnaires (provider and subject) based on the guidance provided by Corrona [41].

Prior to participating in the data collection activities, each site must obtain approval from Corrona and an Independent Review Board. It also has to have a designated Principal Investigator who is the point of contact for all related research activities. There is a broad range of subscribers to the Corrona database, including biopharmaceutical companies, payers, research organizations and others that seek answers to an array of research questions [36, 41].

Data considerations

The data in the Corrona rheumatoid arthritis registry is collected from both physicians and patients at the time of a clinical encounter. This is prospective cohort study where patients and their physicians are followed longitudinally, to generate real world data that captures vital information covering all aspects of the patient care history, such as presentation, diagnostic investigations, treatment, adherence, health outcomes and adverse events, among others [41].
It is a large registry with over 40,000 patients enrolled, translating to over 140,000 patient years of longitudinal follow-up. Within this expansive registry, there is a subset of approximately 10,000 patients that have additional personal information which is linkable to other datasets, subject to the strict privacy considerations as prescribed by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) [41].

**Services**

By carefully analyzing data on patient preferences, provider prescription patterns, trends of adverse events, health outcomes and costs, among others, the Corrona registry provides a useful basis for answering a wide range of medical and safety concerns of decision makers within the biopharmaceutical industry. Meanwhile, payers are better able to better assess the comparative cost effectiveness of products for reimbursement purposes, while regulators are able to realize robust pharmacovigilance services for specific products of interest [41, 42].

Further, it is possible to generate useful insights that are relevant to the commercial considerations of biopharmaceutical clients, such as adherence and persistence patterns, market sizing and penetration, all that are vital in ensuring product success. For those interested in specific studies, including conducting clinical trials, it is possible to leverage data from the Corrona registries to identify optimal sites for study participant recruitment as well as the potential to apply adaptive trial approaches that are likely to be more efficient. Broadly, such insights and possibilities are essential in the efforts to accelerate innovation in order to develop and deliver healthcare products that create value for the patient [41].

In terms of actual care delivery, participating sites have access to innovative tools provided by Corrona that enable physicians to holistically view and track their patients in ways that optimize care. For instance, it is possible to monitor patient adherence and progress to help optimize treatment outcomes. Further, providers in partner sites can securely collaborate with their peers within the network in order to promote joint learning [41, 42].

Further, using data from the Corrona registries, it is possible for researchers to conduct comparative effectiveness analysis, with head to head comparison of different therapies. In addition, it is possible to assess treatment patterns and determine adherence or persistence to treatment as basis
for optimizing treatment outcomes for rheumatoid arthritis patients. The rich dataset also allows for subpopulation analysis, whereby the patient heterogeneity could be interrogated and outcomes assessed, including the those reported by the patients themselves. Safety and adverse events, a crucial consideration throughout the product life cycle could also be tracked using the registry data available through Corrona.

**Business model**

As previously mentioned, the Corrona model for data collection is dependent on having a robust network of data providers in the form of partner clinical sites. The data that are collected from these sites undergo rigorous quality checks for completeness and accuracy, prior to storage in the Corrona data warehouse. From the data warehouse, Corrona experts that comprise of epidemiologists and biostatisticians, access the data for analysis depending on the requests submitted [36, 41].

The Corrona registry services are structured in the form of annual subscription to cover data access, where different clients submit analysis requests that are fulfilled by the analysis team at Corrona. Further, there are additional fees based on consulting time of the data analysis team. The output is normally a manuscript or abstract for publication in a scientific journal. In this case, despite the request coming from a biopharmaceutical company, Corrona maintains independence and rigor in the process by mandating an academic lead as the final arbiter in the entire analysis and publication [41].

Counting the research products from the Corrona registry that have been published, which included manuscripts and abstracts; there were 17 publications, in 2015, 11 publications in 2016 and 12 publications in 2017. These articles were wide ranging and covered topics including, treatment outcomes for specific drugs; comparative drug safety analysis; time to event analysis; treatment switching and their reasons, among other topics [41].

However, it should be emphasized that Corrona registries, is but one of the many examples of data access models that have emerged across the healthcare landscape. In the ensuing sections, we systematically explore the data access considerations that are driving the trends across different
healthcare systems, followed by a description of the emerging models that are serving the biopharmaceutical industry.
Health data access considerations

With the recognition of the value creation opportunities that could be harnessed from systematic data sharing across the healthcare delivery value chain, various stakeholders have come up with different initiatives to drive the process. For instance, some regulatory agencies, are increasingly requiring greater sharing of data by companies seeking to market drugs and devices. In addition, influential organizations operating at both the global and national scale such as the World Health Organization, Institute of Medicine and the National Academy of Medicine have called for responsible sharing of clinical trial data as a way to accelerated innovation and improvements in health outcomes [30, 31].

Many other organizations, including leading drug companies, the European Medicines Agency (EMA), the National Institutes of Health, and the Bill and Melinda Gates Foundation (BMGF), have made it clear that study reporting and data sharing in medical research are imperative [51-53]. Notably, the policy of the BMGF is that the data underlying published results should be made available and open immediately. Organizations that receive research funding from the organization are required to make available the data used upon publication [52]. This gives further credence to the proposal by the International Committee of Medical Journal Editors (ICMJE) to accelerate the transformation to a culture of open science. The ICMJE believes that the custodians of data generated through interventional clinical trials have an ethical obligation to responsibly share the same, within a 6 month period of the publication of the results [54]. In this section, we present insights generated from key informant interviews in terms of the key considerations that determine data access with healthcare ecosystem.

Data demand factors

Study participants indicated that overall, the health ecosystem is witnessing increased demand and supply of data, with the promised to catalyze innovation and improved learning within the healthcare system. Majority of the participants further noted that, the rising healthcare costs and reimbursement trends are raising the bar particularly for health providers and biopharmaceutical companies to clearly demonstrate value for products and services, and hence the demand for data.
It was further clarified that the fee-for-service models of healthcare delivery, where providers were compensated based on the volume of services provided is losing ground to value based models that focus on quality and health outcomes.

Most of the participants also noted that the wave advocating for evidence based medicine and practice across the health system has also emerged as a clear imperative for systematic triangulation of multiple data sources to generate information to guide care decisions. An example provided by a participant from the biopharmaceutical industry is the need to identify specific subpopulations that are most likely to benefit from certain interventions in order to develop medicines that are effective for specific target groups. It was confirmed, by the participant from the health insurance industry, that the need for patient stratification is particularly relevant in the advent of risk sharing models where judicious use of resources and evidence is encouraged.

Participants in addition noted that the demand for data in the healthcare system is reinforced by the need to differentiate or identify the comparative benefit of the various health interventions that are offered to patients, as well as the imperative to ensure safety and prevent harm. More specifically for the biopharmaceutical industry, participants observed that improved access to patient data can optimize the clinical development of products e.g. in the identification of biomarkers and recruitment of appropriate patients to ensure the success of pharmaceutical products.

**Data supply factors**

On the supply side, most participants observed that an interplay of various factors are coming together to accelerate the amount of data available and its utility in the healthcare ecosystem. For instance the participant from the health provider group noted that the rapid adoption of electronic medical record systems by physician practices and hospitals, across the United States, has led to exponential growth, unleashing tremendous amounts of useful data on patterns of care delivery and outcomes. In addition, participants from both the health insurance industry and health provider group suggested that the emergence of health information exchanges among different healthcare providers at regional and local levels has the benefit of pooling large amounts of data to facilitate learning across the broader healthcare system.

Advancements in mobile technology platforms, were also identified by majority of participants as
contributing to the growth in the supply of patient generated data – from both the healthcare system and beyond. It was suggested by a participant from the biopharmaceutical industry, that for example, providers can learn more about a rheumatoid arthritis patient functional status by assessing their mobility and self-reported outcomes from data generated from mobile devices. In addition, payers could also learn about their patient’s financial status, purchasing preferences and lifestyle choices – with the aim of incentivizing preventive health measures.

Other technological advancements that have led to the increased supply of data in the healthcare ecosystem that were identified by various participants, include, the increased to capacity to digitize data from various sources, include motion and images etc. as well as advancements in analytical methods that are able to handle unstructured datasets and derive useful insights. Further, it was observed by the participant from research and academia that the advancements of the many of these technologies, including mobile and diagnostic such as genetic tests, were coupled with declining costs, which facilitated wide adoption.

**Regulatory factors**

Many of those interviewed also felt that the US government was instrumental in providing an enabling regulatory and legislative frameworks to facilitate data sharing among stakeholders in the healthcare ecosystem. Participants from the biopharmaceutical industry cited the passage of the 21st Century Cures Act, which they said had spurred growing interest in the use of real-world data for drug development and regulatory processes.

It was also noted by participants from health insurance industry that, in addition, to the formulation of policies that support data sharing, government was also leading and providing guidance measures to support wide access to data among stakeholders. Many of the participants further identified the role of government as facilitating (e.g. data standardization and creation of portals for easy access) the process through specialized institutions such as the Food and Drug Administration (FDA), and Centers for Disease Control and Prevention (CDC). Further, it was observed that government provides overarching guidance, through HIPAA to guarantee privacy and confidentiality of health data, a necessary prerequisite for responsible data access and utilization. Various government agencies were also identified by participants as directly funding some of the initiatives that promote broader data and information exchange across the healthcare
ecosystem, such as US FDA Sentinel and PICORI which has initiated PCORnet.

It was also observed by the participants from the biopharmaceutical sector, that government agencies in their regulatory role - aimed at approving safe and efficacious health products - also has a direct impact on data access within the healthcare ecosystem. Further, the requirements for better evidence to demonstrate value, as well as adaptive approval mechanisms through which promising products could gain expedited approval, are all factors that have a bearing on data generation and access within the healthcare ecosystem. Overall, the primary goal is to curb rising healthcare costs, improve quality and healthcare outcomes.
Data access models in the biopharmaceutical industry

In the quest to satisfy the emerging demand for real world data and insights, various business models have emerged through which biopharmaceutical companies access data and information. Ranging from direct purchase and subscription models to broad collaborative coalitions, companies are making efforts and first/fast movers in this space are already seeing positive results [21, 25, 55]. Without claiming to be exhaustive, in this section, we summarize some of the insights from participants describing the various avenues or models through which the biopharmaceutical industry accesses data with the healthcare ecosystem. An attempt has been made to classify the various models into distinct categories, but we caution that, these models are rapidly evolving and have different overlapping attributes that make it difficult to fit perfectly into one class or another.

Data vendors

Participants indicated that some firms and institutions that have the capacity to curate data from multiple sources and generating useful insights are commercializing these capabilities and extending access to biopharmaceutical companies. It was observed that many of these data vendors or data aggregators were structured in fee-for-service or subscription based models, where they provide data driven informatics based on the integrated data sets [56]. Participants pointed out that in some cases, the clients of the data aggregator companies are also the one that contribute the data that is used in the analysis to generate the insights.

An example of a data vendor is IQVIA [57], which triangulates multiple data sources including providers and pharmacies to generate broad insights that are useful to biopharmaceutical industry decision makers. The insights generated from the data vendor firms touch on among others, research and development, such as clinical development to optimize results; real world evidence and value demonstration solutions; and commercialization strategies.

Stand-alone analytics divisions within companies

Another data access model identified by participants included the stand-alone analytics division within a parent company. Participants from the health insurance and biopharmaceutical industries...
indicated that this category of data and insight providers usually comprises of those stakeholders that have privileged access to large amounts of data that they could leverage to provide useful insights to decision makers in the biopharmaceutical industry. These include large payer organizations that operate stand-alone analytics divisions. Some of the conspicuous examples in this group include, OptumInsight for United Health, ActiveHealth for Aetna and HealthCore for Wellpoint [21, 56]. From their extensive data bases, these organizations are able to provide robust real world evidence to stakeholders keen to understand various aspects of costs and health outcomes. This is particularly useful in the advent of performance based reimbursement mechanisms.

Participants further observed that this model has seen a further evolution whereby biopharmaceutical companies are collaborating with other stakeholders to answer specific questions or generate more nuanced evidence. For example OptumInsight has a collaborative arrangement with Merck, aimed at publicly sharing analytic insights, findings and recommendations to help inform and facilitate the understanding and use of pragmatic approaches to outcomes-based risk sharing agreements in the health care system. To create complimentary synergies, OptumInsight provides deep expertise in pharmaceutical contracting, health economic and outcomes research, health policy and regulation, as well as access to the data required to fit the new models. Meanwhile, Merck provides data science and outcomes research expertise, a deep understanding of the pharmaceuticals marketplace, and experience working with large private and public customers to create flexible pricing models across multiple therapy areas [58].

Another form of unique collaboration with a stand-alone analytics platform is Optum Labs which brings together Optum of United Health and Mayo Clinic. This platform facilitates the linkage of claims data with clinical health records in a manner that allows for research into cost reduction and improvement of health outcomes. Pfizer is one of the members that have recently joined this collaborative arrangement with Optum Labs to generate real world evidence that is useful to the broader healthcare ecosystem [59].

**Non-profit local data stores**

The local data store is the traditional model whereby the data is hosted by its owner, which is usually a non-profit organization. This could be an academic or research institution; non-profit
agency; private or public organization that is the custodian of a given dataset. In terms of data access, external parties request permission to use datasets for specific analysis. Normally, the data requestor submits a research protocol that is reviewed and if approved, the requestor is allowed to view and analyze the data for a specified period of time. However, in most cases the requestor does not have the rights to download the data.

There are various adaptations of this model, including those that combine some aspects of multi-organizational collaboration. A participant from the biopharmaceutical industry identified, the Yale University Open Data Access (YODA), as one such initiative where an academic institution acts as a data holder and custodian for a pharmaceutical companies, Medtronic and Johnson & Johnson. The YODA partnership is an example that uses a “trusted intermediary” approach, in which an independent partner facilitates the process of data sharing among different stakeholders [62]. The data holders have given the YODA full jurisdiction over the data access considerations as a way to foster transparency and accountability.

**Private-venture backed data aggregators**

Participants from the biopharmaceutical industry noted that more recently, the healthcare ecosystem has seen the emergence of a new kind of data aggregator companies backed by major venture capital firms and partnered with big pharma companies. In the oncology therapeutic area, conspicuous examples of this trend include companies like CancerIQ, Syapse, Foundation Medicine and Flatiron Health. As an example, Flatiron Health which was recently acquired by the pharmaceutical company Roche, works with cancer centers to transform patient electronic health records (EHR) into a continuously updated research-grade database [60]. Through this acquisition, Roche hopes to accelerate industry-wide development and delivery of breakthrough medicines for patients with cancer.

A participant from research and academia, also highlighted Aetion, as another venture capital (New Enterprise Associates) and biopharma (Amgen Ventures) backed data aggregator that works with all healthcare data including claims, electronic health records, registries, and clinical trials, to generate real world evidence. The Aetion evidence platform has federated features in its construct
such that its can work with data wherever it is located [61].

**Federated models**

Considering that meaningful analysis entails triangulation of multiple data sources, the federated query model has increasingly become an attractive model for data access [63]. This facilitates specific data access and analysis leveraging multiple independent databases that are in the custody of different owners that might even be geographically dispersed.

Participants indicated that in a federated model, databases are linked in such a way that they can respond to queries from authorized users as if the data were hosted in a single database. Therefore, it is possible to access the collective power of all the data within the network, without submitting multiple queries, making this model attractive. In addition, the data owners, still retain full control of their data and only share the allowable features. Many participants felt that this was a useful attribute considering the importance of privacy, confidentiality and data security considerations in the health data landscape.

Normally, within the federated network, there is a trusted intermediary (but in some cases this might not be the case) that facilitates the processes related to data access such as query services. Biopharmaceutical companies may purchase such query services from the responsible intermediaries in order to gain useful insights from the data from multiple stakeholders within the network. Examples of federated models, are the US FDA Sentinel Initiative and PCORnet program that combine datasets from various sources for comparative effectiveness and outcomes research [44-46]. Figure 4, is an illustrative presentation of the various stakeholders that are represented in the federated data access models.
Figure 4: Federated data sharing model

Adapted from PricewaterhouseCoopers: Pharma 2020: Challenging business models: Which path will you take?
Illustrative Examples

As previously mentioned there are various data access mechanisms that have emerged to meet the data needs for different stakeholders across the healthcare system. The federated data access model, provides a useful framework that allows various stakeholders to collaboratively share their data. This is essential to answer specific questions that different stakeholders may have, but lack sufficient data to generate useful insights alone.

In this section, we briefly describe some illustrative examples of federated data access models comprising of different stakeholders from across the healthcare landscape in the United States. To represent the wide variety covered by the federated data access models, we briefly describe, the US FDA Sentinel Initiative, The National Patient-Centered Clinical Research Network, (PCORnet) and the Center for Health Information and Analysis (CHIA) of Massachusetts State. In each example, we describe the stakeholders represented, the data considerations, including generation, custody and analysis; services offered and the business model.

In appendix E, we further compare these 3 illustrative examples with the Corrona registry, that was covered previously. This highlights the various differences that are important in the organization of a data access model, such as the type of sponsoring organization, geographical scope, financing and therapeutic area(s) covered. These are important attributes that determine the overall utility and relevance of the data access model to various stakeholders in the healthcare system.
US FDA Sentinel Initiative

The US FDA has the responsibility for advancing public health by ensuring the safety, efficacy and effectiveness of medical products. It is also charged with the mandate to help speed up innovations that make medical products more effective, safer, and affordable; as well as dissemination of critical information related to the same. In order to effectively deliver on this overarching mandate, the FDA Amendment Act was passed in 2007 [43]. This amendment required that the organization work closely with public, research and academic entities as well as the private sector to develop a robust system to use the existing electronic healthcare data from various sources for pharmacovigilance purposes, targeting drugs, vaccines and other biologic products [44].

Key stakeholders

The FDA Sentinel Initiative is an example of a federated model where multiple stakeholders collaborate to advance the evidence base, which is not feasible in isolation. Within this structure, there is a coordinating center, which is led by the Harvard Pilgrim Health Care Institute (HPHCI), and includes other partners drawn from various sectors that contribute both healthcare data and technical expertise. The sentinel coordinating center is comprised of the Sentinel Operations Center (SOC), and the advisory groups that are responsible for the overall implementation of the collaborative [44, 45].

Without being exhaustive, some of the organizations represented in the FDA Sentinel Initiative include, private data analytics companies such as Aetna analytics, IQVIA and Optum; payer organizations such as Humana, Blue Cross Blue Shield of Massachusetts; academic institutions, such as Rutgers, Harvard University, Alabama University, Duke School of Medicine, among others. Hospital groups, are also represented in this federated structure, and notable ones include, Kaiser Permanente (and its research affiliates) and Hospital Corporation of America, among others [45].
Data considerations

Through the federated data infrastructure comprised of multiple partners, the FDA Sentinel has a vast and secure access to electronic healthcare data records from approximately 200 million patients. The secure infrastructure and data governance procedures employed guarantees patient privacy and confidentiality in accordance with the HIPAA guidance [44-46].

The partner institutions and the FDA are responsible for the custody and stewardship of the individual datasets in their possession, without the requirement for physical transfers. Partner organizations have in their possession various data sources that are generated through their regular activities, such as electronic health records, claims and billing data, registry data and pharmaceutical product consumption patterns. These types of data are referred to as original source data, in contrast to external source data that is collected outside the routine activities of the partner organization. This may include specific studies to confirm exposures or outcomes of interest, e.g. adverse events resulting from immunization activities [45, 46].

In order to conduct data analysis, the SOC distributes standardized data queries that are implemented by respective data partners and the results shared back to the SOC for collation and onward sharing with the FDA. Unless explicitly specified, data partners provide the data in summary form, without individual identifiers. Non-summarized data is only available to authorized personnel within the SOC. For the analysis process to be effective, the data partners maintain and provide access to data employing a commonly agreed data model across the wider partnership [45].

Services

The US FDA Sentinel Initiative is meant to enhance the ability of the FDA to identify and assess medical product safety issues in the effort to advance public health. This system allows for effective monitoring of the safety of regulated medical products, which is vital for multiple stakeholders within the healthcare ecosystem [44, 45].

Apart from the regulatory function, the pharmacovigilance information and other insights generated from the Sentinel Initiative are essential to in aiding biopharmaceutical companies learn
about the performance of their products, and possibly their competitive market place. Overall this interaction is vital in supporting efforts around product development and access in order to advance public health.

**Business model**

This is a federally funded initiative. US FDA awarded HPHCI a contract for the Mini-Sentinel pilot in 2008. HPHCI has since developed a federated data network to meet the set objectives. HPHCI serves as the main contractor for the initiative and is responsible for all aspects of the distributed network, including secure hosting and maintenance and development of analytic tools and query capabilities, as well as sending queries to partners to answer FDA safety questions [45, 46].

**National Patient-Centered Clinical Research Network**

The National Patient-Centered Clinical Research Network, PCORnet, is another innovative federated data sharing network that comprises of multiple stakeholders with the aim of enabling faster, easier and less costly clinical research that is informative to decision making. This overarching objective is to be achieved by harnessing the power of large amounts of data through partnership with various health system stakeholders including patients. PCORnet is an initiative of the Patient-Centered Outcomes Research Institute (PCORI), which is an organization that was established by government to fund research that can help patients and their care-givers make better-informed decisions about the healthcare choices [47].

**Key stakeholders**

PCORnet is a large nationally representative network made up of groups of partner networks comprising of various kinds of stakeholders including, clinical service providers, patient groups and health plan research organizations. In total, there are 13 clinical data research networks, 20 patient backed research networks and 2 health plan research networks. Some of the patient groups involved include, chronic conditions such as arthritis, multiple sclerosis and Alzheimer’s disease, among others [47].
Within this federated structure, the PCORnet coordinating center, is led by the Duke Clinical Research Institute, Genetic Alliance, and HPHCI that provides technical and operational support to the partner networks [48]. Having a neutral party (usually, an academic institution) as a coordinating center is a common feature for federated data collaborative mechanisms.

Data considerations

Through its vast network, PCORnet provides access to vast amounts of data that is collected routinely in real world healthcare settings through the continuum of care, including hospitals, clinics and physician practices. Similar to other federated models of data access, each partner network enjoys the custody and stewardship of its own data. These data are collected and stored following a standardized format or a common data model within their own institutions [47, 48].

PCORnet has an online portal through which the request for data by individual researchers are submitted following established protocols. These requests are then reviewed by the coordinating center, which launches standardized queries on the partner data sets (without transfer from their own institutions) to answer the requested research questions and summary results are sent back to the original requestors.

Services

PCORnet uniquely empowers patients, care-givers and other health stakeholders to be engaged in the research process that promotes learning across the system and advances public health. Furthermore, the resources available through PCORnet could be leveraged by researchers in various types of research including exploratory assessments, observational and interventional studies. The PCORnet robust infrastructure, offers researchers the ability to query a powerful pool of patient data which is of national scale to accelerate the generation of useful insights to facilitate learning and decision making among different stakeholders. Priority is given to patient-centered research that has the capacity to help people make better decisions [47, 48].

Business model

PCORnet is an initiative by PCORI, which is government-sponsored organization charged with investigating the relative effectiveness of various medical treatments. PCORI funds a wide range
of organizations, including those within the PCORnet network through grants to conduct various types of research that is patient centered and informs care decision making [48].

**Center for Health Information and Analysis**

The Center for Health Information and Analysis (CHIA) is an independent state agency established in the state of Massachusetts in 2012 to act as the primary hub for health care data and a primary source of health information that support policy development. In addition to its role as a data steward and custodian of sensitive health data, the agency is tasked with the responsibility to develop robust analytic systems and report on different aspects of healthcare such as quality, affordability, utilization, population access and outcomes within the state. The overarching mission of CHIA is to prompt a transparent healthcare system and empower the stakeholders thereof to make informed and evidence based decisions [49].

**Key stakeholders**

CHIA as the primary informational and analysis hub within the Massachusetts health care system, has close connections with other stakeholders, including policymakers and state agencies, public and private payers, providers of various types, employers, researchers, and the residents of Massachusetts. It collects and maintains large detailed datasets from various stakeholders operating within the health system [49, 50].

Many state agencies use the healthcare data in the custody of CHIA for their research, regulatory activity, and operations. Considering that their needs are diverse, CHIA facilitates this process by having a unified informational and analytics hub, in order to minimize overlapping and complex data requests from different agencies within the state [49].

**Data considerations**

CHIA is mandated to collect data from health care payers, providers, provider organizations and third-party administrators within the state of Massachusetts, in order to promote transparency and access in all aspects of healthcare delivery. Through its analytics, CHIA generates reports on healthcare utilization, performance of hospitals, insurance coverage among others [49].
Some of the key data sources that are collected by CHIA include, the Massachusetts All-Payer Claims Database (MA APCD), Acute Hospital Case Mix Databases (AHCMD), and Massachusetts Hospital Financial Performance information. The MA APCD mainly covers information from both public and private payers including, claims of all types, eligibility of members and benefit design. This database also provides information on the population coverage by different health providers. The AHCMD comprises of acute care patient level data from hospitals across Massachusetts. This broader database further comprises of constituent databases on inpatient discharge, outpatient observation and emergency department. When carefully analyzed these databases provide useful information to the healthcare system which includes utilization patterns, readmissions, hospital market analyses and comparative charge analyses [49, 50].

In addition to care delivery activities, CHIA collects and maintains large detailed datasets from hospitals and other health care providers as well as payers across the state, covering different aspects of expenses, payments and overall costs of care within the healthcare system [50].

Services

By law, CHIA is mandated by the state of Massachusetts to collect, analyze and disseminate health care information to assist in the formulation of health care policy and in the provision and purchase of health care services. The reports generated from CHIA cover useful aspects such as healthcare costs, health outcomes and quality information, as well as the performance of health system. For instance, hospital financial performance is routinely reported through annual and quarterly reports, which cover individual and aggregate data across a spectrum of health facilities. Overall, CHIA offers differentiated analytic tools using the data in its custody to foster transparency, accelerate cost containment, and quality improvement within the broader health care system [49, 50].

Business model

CHIA is a statutory body that is supported by the state of Massachusetts. Some of the data it collects is publicly available, while other datasets are only available by application. Researchers and others working in the public interest can access these datasets in order to generate additional
insights that advance public health. Government data requesters have designated pathways to access data within CHIA in order to perform their functions. Non-governmental requestors are also able to access certain datasets with payment or waiver of certain administrative fees [49].
Conditions for success and potential risks

Base on the literature review and expert insights from key opinion leaders, it was clear that there are a myriad of factors that determine the success or failure of data access models within a healthcare ecosystem. Therefore, a clear identification of those factors is vital for any efforts towards long-term success in harnessing the promised benefits of improved data access. In this subsection, we highlight some of the key determinants of success and barriers to progress in data access within the healthcare ecosystem.

Determinants of success

Given the diversity across the broader healthcare ecosystem, most of the participants interviewed were in agreement that building trust among stakeholders is a priority step for any meaningful efforts towards developing sustainable data access models that drive innovation and ultimately improve the patient health outcomes. This is particularly true given the wide range of stakeholders in the healthcare space with multiple competing priorities and interests. It was emphasized that in some federated models, trust among stakeholders is achieved by having a “trusted intermediary” that coordinates and facilitates the activities of the broader collaborative. Further, the trust factor covers the privacy and confidentiality considerations (for both patients and organizations) that must be taken into account [23, 48, 63].

Participants further indicated that establishment of strong governance systems and ground rules from the outset is essential for long term success of any data access initiative that involves a number of stakeholders. Other participants further suggested that in crafting the governance systems, efforts should be made to ensure that governance structures are representative in order to ensure stakeholder engagement and commitment. It was cautioned that, stakeholders that feel disadvantaged in a data collaboration arrangement are likely to lose interest and be disengaged and vice versa.

Majority of participants observed that, in as much as the healthcare system represents various stakeholders, alignment of objectives among stakeholders is essential. It was suggested that by beginning with the primary healthcare system goal that places the patient at the center and seeks
to improve and restore health, various stakeholders are able to gain consensus on how their individual objectives align with one another. For example, the value health providers generate depends on the revenues payers raise and the medicines biopharmaceutical companies make. Further, the value biopharmaceutical companies generate depends on getting access to the patients whom providers serve and income from the payers who fund those providers. This interconnectedness in value creation among stakeholders could be expanded to explore opportunities for sustainable win-win outcomes for all stakeholders to collaborate rather than pursuing narrow and short-term interests [20, 24, 31].

Operationally, participants suggested that having a common “language” of communication among different stakeholders is essential to ensure success. According to the literature and participant insight, the common language is achieved through comprehensive efforts to standardize data among partners to facilitate ease of sharing and analysis. In fact, participants indicated that many of the federated data access models described above have common data models that allows for standardized and faster analysis by deployment of standardized queries.

Organizational and platform adaptability and leveraging technology to expand access is another attribute of successful data access models that was identified by various participants. It was observed that stakeholders, particularly large payers and biopharmaceutical companies have heavily invested in data analytics platforms that have the capacity to generate useful insights to drive innovation and improve health. However, these expensive investments would only realize their ultimate potential if they are agile and adaptable enough to shift with the rapidly changing landscape.

Further, participants suggested that for the data access models to be credible and gain trust of decision makers, they should strive to maintain high standards of scientific rigor and independence from undue influence when generating and disseminating evidence. For many, it was clear that data is emerging as a big currency across various industries, but funding for initiatives to harness the power of data remains low. Therefore, funding considerations are essential in pursuing this quest. In particular, the many stakeholders that gain tremendously from the data insights generated should be encouraged to reinvest in this space to create a virtuous cycle.
Barriers to progress

Some of the obstacles to progress in data access that were identified by participants include resistance to change- as a result of organizational inertia, legacy systems or individual level factors. Considering individual level incentives, participants observed that normally, health providers are accustomed to making decisions independently and collaboration needs a significant mind shift. Further, participants noted that in some cases, stakeholders have made large investments in legacy systems that are difficult to standardize or allow interoperability with other stakeholder data bases. This creates inefficiencies and inertia that hinders any steps towards broader access to data among different stakeholders [14, 18, 25].

Furthermore, it was observed by participants that in most cases data is often collected and analyzed in siloes among different healthcare stakeholders. In fact, even within a single stakeholder, it was observed that the silo effect is within groups and departments further fragmenting data and making it difficult to generate useful insights. In other cases, there is been underinvestment in data and information generation by stakeholders leading to a paucity of evidence in decision making. In this scenario, the limited use of evidence, leads to further underinvestment in data and analytics, effectively establishing a vicious cycle that curtails all aspects of data utilization within the healthcare system.

Majority of participants also identified poor data quality as another fundamental challenge that hinders progress. In fact, in cases where data is not used for decision making, there are little incentives to improve its quality and hence establishing a vicious cycle of limited use and poor quality of data. Further, lack of standardization of data collection procedures and quality constraints also forces researchers to go through difficult efforts to address gaps and inconsistencies in data through methods that may not yet have wide acceptance for statistical validity. This in turn has a negative impact on the utility of the information that is generated and thus curtailing efforts to improve data access [14, 56].

Cost is another important barrier to the data sharing efforts that was identified by many participants. It was noted that on a large scale, most of the processes linked to data collection
analysis and dissemination are expensive and beyond the reach of many healthcare system
stakeholders. For instance, small payers are not able to invest in sophisticated data analytics
platforms, while small healthcare providers might not have the capacity to install the state of the
art electronic medical record system. It was noted that unless carefully addressed, cost barriers are
likely to lead to the exclusion of a significant section of stakeholders resulting in incomplete
insights that miss out on large sections of the population.

Some participants also pointed to regulatory hurdles as having a negative impact on data access in
some cases. The need for regulation to be regularly updated to reflect the current realities within
the healthcare ecosystem is imperative. Lack of procedures of integrating data and communicating
information in systematic way among stakeholders could also curtail progress in promoting data
access [14, 18, 25, 56].

Despite the many merits that have been highlighted in this section supporting improved data
access, it is also useful to be aware of its potential risks in order to mitigate them accordingly.
Personal health data is often highly confidential, and therefore any privacy or confidentiality
breeches, allowing unauthorized access and use of this data could lead to harmful repercussions
across the health system. For example, leveraging data to identify high risk subpopulations in order
to adversely select them in private health insurance coverage constitutes harmful use of data.
Others, my include using a privileged position in data access to effectively establish unfair
competition and healthcare cost escalation. Therefore, data sharing does set the stage for
considerable ethical risks if not appropriately governed.
Discussion

There is a clear value proposition that improving access to data across the healthcare ecosystem would be beneficial in driving innovation and improving patient outcomes [64]. Figure 5, is a schematic presentation of a basic data integration framework among key stakeholders, leading to enhanced innovation potential and opportunities to improve health. Therefore, like other stakeholders in the healthcare ecosystem, the biopharmaceutical industry stands to realize tremendous gains in a regime of improved data access [20-25]. For instance, collaborating much more closely with other stakeholders in the healthcare sector will enable the biopharmaceutical industry to expand the utility of its products and interventions by gaining deep insights into the patient healthcare seeking experience and interaction with the broader healthcare system [64].

Further, such collaborations in learning the patient experience presents additional opportunities for closer alignment in value creation and delivery, among different stakeholders, inclusive of payers, biopharmaceutical industry and providers. This is particularly relevant considering the diversity of incentives and often conflicting interests and competition among different stakeholders. By closely examining data from multiple sources, it makes it possible to identify areas of intersection where partnership with often competing stakeholders could lead to a win-win outcome [20, 22, 23].

Lessons from other fields of science point towards the great benefits of data sharing and collaboration among stakeholder for the common good. There are examples of many physicists, astronomers, and geneticists, who actively shared data from notables projects such as the Large Hadron Collider, the Hubble Space Telescope, and the Human Genome Project. The defining distinction of these successful and landmark undertakings has mainly been centered around having a unifying objective or task to accomplish [65]. In the case of healthcare system stakeholders, this singular objective would be to improve and restore patient health [1, 2]. From this common objective, individual stakeholders, such as payers or providers are then able to derive or determine their specific contribution to the collective success in terms of improvement of patient health.

These considerations are clearly reflected by the key questions, decisions and actions that proactive decision makers within specific stakeholder groups have to make to realize the common objective of improving patient health. For instance, patients would need to understand what are the appropriate life style choices to espouse in order to attain good health. Health providers would need to understand the health interventions that work for different population subgroups for
targeted delivery as well as measures to reduce the escalating healthcare costs. Payers, on the other hand, need to have clarity on the cost effectiveness and cost efficiency of different products and services that they are reimbursing, so that patients benefit.

Clearly, within the biopharmaceutical industry there are many opportunities to discover and create value in the various stages of drug development, by leveraging the power of data sharing across the healthcare system [19-25, 64]. For instance, the biopharmaceutical industry could contribute to the common objective of improving and restoring patient health, by accelerating novel and effective treatment options, when there is clarity on patient biomarkers, comorbidities and other useful characteristics for specific subpopulations. Obviously, by examining multiple data sources, useful and informative insights could be generated to guide the way clinical trials are designed and implemented, with the aim of achieving efficiency and cost effectiveness. Patient recruitment and follow up could also be optimized by carefully examining longitudinal data bases derived from routine care and other research activities, providing further opportunities for cost savings and innovation [21, 23, 25]. In addition, data from real world settings could be used to demonstrate value in order to gain market access, reimbursement and ensure commercial success of pharmaceutical products.

Therefore, the ongoing digital revolution that has touched many sectors of the economy will definitely have a transformative effect on the way healthcare is delivered. The pioneering organizations that are at the forefront in creating mechanisms to harness the power of data stand to gain immensely. Health providers will be able to understand care patterns that optimize health outcomes for patients and reduce healthcare costs. Further, through nuanced evidence generated from multiple data sources, payers would have the confidence and better insights to select and reimburse only those pharmaceutical products and services that have demonstrable value to the patient.

However, given the complexities of the healthcare ecosystem, there is need to first understand the push and pull factors that are driving the trends, in order to tailor appropriate mechanisms to realize these gains. Factors such as the shift to value based care and evidence based medicine, rather than the volume based healthcare delivery are fundamental in driving the demand for data access across the healthcare system [15-17]. Technology advancements, particularly in the digital and mobile space as well as the proactive regulatory measures are also crucial in enabling this trend towards
increased demand and use of data for decision making within the healthcare system [20, 23, 25].

Collaboration among different stakeholders in the on-going data revolution becomes imperative in the quest to realize the common objective of improving patient health. Thus, understanding the incentives, motivations and barriers to effective data sharing across the healthcare system would be useful in the efforts to structure appropriate governance and operational systems that drive success. Notably, the rapidly evolving legislative and regulatory landscape, the desire to curb rising health care costs, demand for quality care, patient awareness and empowerment are some of those factors that are driving demand for improved data access [21, 25, 26].

To cultivate useful data access mechanisms among different stakeholders, trust is a fundamental consideration that is essential is the efforts towards meaningful engagement. In addition, harmonization and standardization of data standards is vital to ensure quality of data and meaningful analysis to generate evidence. Transparency and inclusivity among different actors is also another important consideration in the effort to cultivate useful data sharing models. Some of the federated data access models highlighted above are an attempt to address these considerations by having a “trusted intermediary” organization, that coordinates the various operational activities on behalf of the partners in the collaborative [46, 47, 63].

In interpreting the findings of this study, we are cognizant of its potential limitations. The topic “data access” in the health care space is broad and there is an expansive amount of literature that we have not been able to consult given time and resource limitations. Further we have based our analysis on a limited set of industry experts that we could conveniently sample. However, we have mitigated for these limitations by ensuring that the selected participants, comprised of a diverse group drawn from different stakeholders of the health system such that the discussion was rich and informative. In addition, qualitative studies offer the benefit of a deeper investigation into important policy matters and perceptions that could be concealed through simple aggregation methods.
Figure 5: Data sharing among different stakeholders in the healthcare system

Adapted from: McKinsey & Company. The "big data" revolution in healthcare: accelerating value and innovation. 2013
Conclusion
The growing interest in data sharing will not translate into progress without the development of supportive infrastructure and policies that enable data to be shared effectively and responsibly among stakeholders. Evidence suggests that various types of collaborations among stakeholders are likely to be successful when objectives and incentives are aligned [1, 22, 23]. By focusing on the common objective of improving and restoring patient health, different stakeholders are likely to identify where their data contribution makes a positive difference to enhance health outcomes.

Overall, depending on the specific focus of the collaboration, those that catalyze or initiate data sharing models, should seek to adopt flexible models that would allow for growth as other stakeholders get coopted. In the face of a rapidly changing regulatory ecosystem and technological advancements giving rise to new information demands, agility and adaptability is essential to ensure long-term relevance and sustainability of the data access models [30, 31]. The distributed or federated model of data access where stakeholders retain custody and stewardship of their own data, while effectively contributing to collective data sharing and research through a coordinating intermediary, have shown promise as scalable options within the healthcare ecosystem [47, 63].

Overall, data access should not be seen as a destination but as a process that is to be continually improved and adapted to meet the needs and demands of a rapidly changing healthcare environment, which is intimately anchored on the primary objective of improving patient health. Therefore, organizations that would gain the most from data are those that stay grounded on these principles and focus on the patient.
Reference


44. US FDA. FDA’s Sentinel Initiative [internet]. 2014; Available at: https://www.fda.gov/Safety/FDAsSentinelInitiative/ucm2007250.htm : Accessed on 7th May 2018


47. PCORnet. PCORnet, the National Patient-Centered Clinical Research Network [internet]. 2018; Available at: http://www.pcornet.org/ . Accessed on 6th May 2018


50. Center for Health Information and Analysis. Overview of the Massachusetts All-payer claims database. CHIA. 2014. Available at: http://www.chiamass.gov/ma-apcd/


64. MIT Center for Biomedical Innovation. NEWDIGS – Catalyzing the Transformation of Healthcare Innovation. 2010. Available at: https://newdigs.mit.edu/catalyzing-transformation-healthcare-innovation

Appendix A: Clinical picture of rheumatoid arthritis

The clinical presentation of rheumatoid arthritis varies from patient to patient. However, most patients commonly report an insidious onset of pain with symmetric swelling of small joints—typically involving hands or feet. Morning stiffness in and around the joints, lasting at least 1 hour before achieving maximal joint flexibility, is a typical sign of the disease. However, in some cases, the disease may affect other joints (including the large ones and without a symmetrical pattern) as well as present with systemic symptoms such as general body weakness and fatigue.

The disease often runs a chronic course, with varying severity over time. As shown in Fig.6, a confluence of genetic and environmental factors act to trigger an autoimmune response that leads to joint inflammation. This stage is characterized by joint pain and swelling, with minimal joint destruction if any.

Progressively, without adequate control of the disease’s activity, the chronic inflammation invariably leads to the development of various degrees of joint destruction, coupled with other systemic comorbidities. These include serious cardiac and pulmonary complications that could be fatal. Rheumatoid arthritis patients have also been identified to be at high risk of developing debilitating conditions such as osteoporosis, increased risk to infections and malignancies, particularly lymphoma.

---

Figure 6: The clinical course of rheumatoid arthritis

Appendix B: General management principles for rheumatoid arthritis

Over the last two decades, there has been significant progress towards effective management of rheumatoid arthritis. This is anchored on early and consistent reduction of inflammation through aggressive treatment leading to minimal joint damage. In addition, the ability to target and customize effective treatment options for individual patients, given the expanded understanding of the disease pathophysiology, has witnessed significant improvements in patient outcomes.

Fig. 7 shows the management strategies for rheumatoid arthritis at different stages of its progression. Disease-modifying anti-rheumatic drugs (DMARDs), form the backbone of rheumatoid arthritis therapy. These drugs act to reduce joint and systemic inflammation and improve function. The leading DMARD is methotrexate, which can be combined with other drugs of this type such as sulfasalazine, hydroxychloroquine, leflunomide, and corticosteroids to great effect.

New biological agents have also been developed as therapeutic agents for rheumatoid arthritis. These are normally reserved for use when arthritis is uncontrolled or toxic effects arise with DMARDs. Prohibitive costs and increased risk of infection are some of the factors that restrict the prescription of biological agents.

Other non-therapeutic measures in the management of rheumatoid arthritis include, smoking cessation, exercise and weight loss, as both primary and secondary preventative measures. In addition, the need to prevent and manage other related comorbidities that may arise is vital in ensuring better patient outcomes.

Therefore, considering the chronic course of the disease and complexity of its management at various stages of progression, rheumatoid arthritis care within the health system requires an evidence based integrated care approach to ensure optimal outcomes. By carefully and holistically examining multiple data streams concerning the patient care journey, decision makers would be empowered to assess progress and take timely corrective action whenever needed.
Figure 7: Prevention and management of rheumatoid arthritis

Appendix C: Patient journey, bad outcomes

Figure 8: Patient journey with poor health outcomes

Adapted from: The National Rheumatoid Arthritis Society Mapping Project: Mapping the Patient’s Journey in Rheumatoid Arthritis
Appendix D: Patient journey, good outcomes

Name: Jane W.
Date of birth: February 1986
Works as a math teacher at a local school
Heavy to moderate drinker and slightly overweight
Smokes 1-2 packs of cigarettes per day

Integrated EHR allows for identification of risk factors such as smoking and avoids duplication of laboratory tests which could lead to cost escalation

Integrated data sharing among providers allows for faster diagnosis and treatment optimization among different providers.

Integration of claims data and patient care and outcomes data to optimize treatment

Patient generates data on lifestyle and mobility innovation collected and used to monitor risk factors, adherence to treatment and functional outcomes

Figure 9: Patient journey with good health outcomes

Adapted from: The National Rheumatoid Arthritis Society Mapping Project: Mapping the Patient’s Journey in Rheumatoid Arthritis
Appendix E: Comparison of data access case studies

The table 4 below highlights some of the different characteristics of 4 data access models that might determine the broader utility to different stakeholders in the healthcare system. These factors include, among others, the types of data shared within the access model, geographical scope, sponsoring organization as well as the type of arrangements available for different stakeholders to share data and analytical insights.

For example, Corrona registry is largely focused on data from health providers that cover rheumatology, while CHIA covers various components of the health system include, claims data that is vital in determining costs of healthcare delivery. CHIA is narrowly covering the state of Massachusetts, while the PCORnet and US FDA Sentinel is nationally representative.

<table>
<thead>
<tr>
<th></th>
<th>Corrona Registry</th>
<th>US FDA Sentinel</th>
<th>PCORnet</th>
<th>Center for Health Information and Analysis (CHIA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combines multiple types of data from across different health system stakeholders</td>
<td>x</td>
<td>v</td>
<td>v</td>
<td>v</td>
</tr>
<tr>
<td>Broad geographical scale (more than 1 state)</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>x</td>
</tr>
<tr>
<td>Sponsored by government /government agency</td>
<td>x</td>
<td>v</td>
<td>v</td>
<td>v</td>
</tr>
<tr>
<td>Covers multiple therapeutic areas</td>
<td>x</td>
<td>v</td>
<td>v</td>
<td>v</td>
</tr>
<tr>
<td>Limited financial barriers to join the data access mechanism</td>
<td>x</td>
<td>v</td>
<td>v</td>
<td>v</td>
</tr>
<tr>
<td>Data partners retain custody of their data remotely without need for physical transfers</td>
<td>x</td>
<td>v</td>
<td>v</td>
<td>x</td>
</tr>
<tr>
<td>Distributed analysis platform</td>
<td>x</td>
<td>v</td>
<td>v</td>
<td>x</td>
</tr>
</tbody>
</table>