IQVIA INSTITUTE FOR HUMAN DATA SCIENCE

Modernizing the Natural History of Disease Research

IQVIA Perspectives from Human Data Science Lab

The term natural history of disease refers to the progression of a disease process in an individual over time, in the absence of any treatment or intervention, according the Centers for Disease Control and Prevention.[1] Being able to not just understand, but also accurately and holistically track the natural history of a disease is fundamentally important in order to assess the efficacy and safety of preventive or therapeutic intervention. In other words, without understanding the course and pathways of disease, it is very difficult, if not impossible, to define the endpoints of a novel therapy that can be evaluated through clinical research. And form the basis for safe and effective therapies.

This challenge is clearest in the realm of rare diseases. Of the 7,000 diseases that affect more than 350 million people worldwide, only a couple of hundred have a known natural history of disease and only 5% a matching FDA-approved therapy.[2] But the discussion also reaches into the need to better understand the natural history of diseases, such as, cancers and neurodegenerative disorders.

Currently, of course, it is the gap in knowledge of SARS CoV-2 and COVID-19 that has pushed the impact of natural history of disease into a harsh and urgent spotlight; here the global research community and the world at large are still left with visibly open, unresolved questions about this perplexing new disease, with dire consequences.

Furthermore, this topic is timely given the rising power of Human Data Science, i.e., the integration of expertise in human science with breakthroughs in data science and technology to advance our understanding of human health, and enable everyone to make better, more insightful decisions.

To discuss the challenges around our evolving understanding of the natural history of disease, with a specific focus on cancers and neurodegenerative diseases, the IQVIA Institute for Human Data Science brought together a multidisciplinary panel of experts from various fields of academic research and medicine, including oncology, neuroscience, rare diseases, epidemiology and reimbursement.

Recognizing the many unresolved questions related to COVID-19 as well as the fundamental uncertainties across many major disease areas, it is now time to modernize the study of the natural history of disease as a centerpiece in medical research and development.

This event was the inaugural session for the Human Data Science Lab, an interactive, open discussion of game-changing topics in medicine, academic research and health policy, designed to advance the understanding of challenging issues and explore new areas for research and solutions in human health.

The first lab session generated a lively and inspirational discussion of the complex and multifaceted topic regarding the natural history of disease, as illustrated in the proceedings from the event: *"Evolving the Understanding of the Natural History of Disease: Perspectives across COVID-19, Cancers, and Neurodegenerative Diseases".*

Inspired by the first Human Data Science Lab, there appears to be a number of important opportunities for further advancing the study of natural history of disease through a variety of academic research endeavors, all of which can be pursued in potential collaborations between academic researchers, other healthcare stakeholders, and the IQVIA Institute for Human Data Science.

1. Advancing natural history of disease studies

While natural history of disease studies have been foundational elements in medical research for decades, there is an urgency and an opportunity to modernize this discipline.

It is urgent because there is a need to better understand complex diseases, such as COVID-19, cancers and neurodegenerative disorders, as gaps in our understanding of the course of such diseases have been revealed with the growing volume of research, and the evolution of genomic and advanced molecular diagnostics; it is the paradox of simultaneously gathering more knowledge, but more unresolved questions.

For these three therapy areas, the challenge and the opportunity converge in our rising understanding – and uncertainty – of prodromal and/or asymptomatic diseases. More and more we are realizing that effective preventative or therapeutic action needs to address the onset of pathology, before there are clinical manifestations and patient symptoms. It is also increasingly clear how the progression of disease is impacted by multifactorial dimensions across biology, genetics, and social, cultural and environmental factors.

Fortunately, there are more opportunities than ever for accelerating the mapping of complex disease progression due to the acceleration of real world evidence and new digitally enhanced technologies, such as predictive analytics with Artificial Intelligence (AI), Machine Learning (ML) and Natural Language Processing (NLP).

There are several areas where real world evidence combined with natural history of disease studies are useful in clinical development. The natural history of disease studies are important, in part, to understand the disease itself and as a strategy to create a more collaborative environment and relationship with clinicians and trial sites. Specifically, they are useful prior to Phase I studies to help inform the primary study objective and determine the key clinical endpoints. Natural history of disease studies can also serve as a potential historical control or external comparator to the single-arm studies that are evaluating treatment.

Predictive analytics, such as AI and ML, can be useful in finding undiagnosed patients. As an example, ML can help identify complex clinical patterns for early diagnosis of disease by leveraging the digital footprint of diagnosed patients to build an algorithm that can identify unique patterns of the disease in patient's pre-diagnosis medical history. In clinical development, AI can help improve patient safety and accelerate results by enhancing operational efficiencies in study design, trial site identification and patient recruitment, pharmacovigilance, clinical monitoring and patient care.

The challenge and the opportunity converge in our rising understanding – and uncertainty – of prodromal and/or asymptomatic diseases.

2. Advancing the quality of large-scale disease registries

Disease registries are critically important sources of information as they generate insights about the evolution of disease in the real world, and enhance our ability to draw comparisons and distinctions across population segments and varying cultural, genetic and environmental vantage points. Emerging large-scale disease registries allow for even broader comparative research across diverse geographic, economic, cultural and ethnic population segments.

Enabling comparisons of diverse large-scale disease registries requires a consolidated methodology to curate, enrich, and create analytic datasets from all of the data generated across various national and international registries and studies, protocols, instruments, cohorts, and time-points. International registries can add scale, but pose additional challenges on two fronts: divergent

methods and types of data collection across national systems, and divergent and onerous privacy regulations, some of which deliberately restrict movement of data across national boundaries.

Overcoming these challenges is worth the effort where scale is the only way to address research questions. For example, natural-history-of-disease studies of ultra-rare diseases can only yield adequate statistical power with sufficiently large populations, and may thus require national or international reach.

Modern registry platforms can usually support a variety of research methods (in addition to natural-history studies), as well as a range of data ingestion methods and data curation processes. Data ingestion methods include both the established manual data capture from clinical sites via case report forms (CRF) or electronic CRF data entry, to more innovative methods such as direct automated data pulls from clinical electronic medical records (EMR) systems, direct-to-patient data collection of patient reported outcomes (PROs) and surveys, and sensor data from outside the clinical setting (e.g., FitBit). A registry platform should support data linking across multiple data ingestion modalities, as well as longitudinal accumulation of data linked to (usually identified) individuals. A registry platform should also support a variety of data curation methods that can be applied to generate data sets suitable for different research purposes, making them an increasingly valuable tool for evidence generation.

Emerging large-scale disease registries allow for even broader comparative research across diverse geographic, economic, cultural and ethnic population segments.

3. Developing a new "Framingham Study"

The Framingham Heart Study was a landmark trial launched in 1948, under the direction of the National Heart, Lung and Blood Institute, with the goal to identify common factors or characteristics that contribute to cardiovascular disease (CVD).

The cohorts of the Framingham study populations were extended over time to encompass more than 14,000 people over three generations. Over the years, careful monitoring of the Framingham Study population has led to the identification of major CVD risk factors, as well as valuable information about the effects of these risk factors such as blood pressure, blood triglyceride and cholesterol levels, age, gender and psychosocial issues. Risk factors such as dementia have also been investigated and are still being studied. Furthermore, the relationship between physical traits and generic patterns have been investigated. The Framingham Study also led to the development of the influential and widely used Framingham Risk Score for Hard Coronary Heart Disease, which estimates 10-year risk of heart attack.

Today, there is a compelling opportunity for designing a new, modern "Framingham Study" with a population mix that is socially, ethnically and culturally more diverse than the original cohort while simultaneously drawing from the exceptional epidemiological power of the original longitudinal, population-based study format.

Potential areas for a new, longitudinal population-based study could be a cluster of diseases around mental health, related co-morbidities and social determinant factors. Alternatively, a longitudinal population-based study could investigate the etiology and course of diseases related to COVID-19 looking at the origins of the pathogen SARS CoV-2 as well as the long-term sequelae of COVID-19, the long-term complications from the disease on multiple organs, and the safety and efficacy of preventative and therapeutic treatment options.

Today, there is a compelling opportunity for designing a new, modern "Framingham Study" with a population mix that is socially, ethnically and culturally more diverse.

4. Enhancing the understanding of the complex interactions of social determinants and their dramatic effect on health inequality

The discussion about the important role of social determinants of health has generated new traction during the COVID-19 pandemic due to the evidence of social and ethnic disparities of the burden of the disease.

The health impact of these factors - whether social, economic, cultural, educational, or ethnic - are highly complex as they are also interwoven with genetics, variations in types of job exposure and the impact of different treatment interventions and the quality of health system interventions and services. What is clear, though, is that there are dramatic – and life threatening – disparities in health outcomes across different populations.

There is an urgent need for better studies that combine epidemiology, clinical, genetic, psychology, data science, sociology, anthropology and behavioral sciences in order to achieve a better understanding of how these various factors intersect and what drivers are representing causality vs. correlation.

This also calls for a rigorous review of methodologies to correctly capture evidence from highly diverse datasets and observations, weed out bias and enhance evidence-based interpretations of research findings.

What is clear is that there are dramatic – and life threatening – disparities in health outcomes across different populations.

5. Improving the quality of data and integration of data

Efforts to improve the quality of data and integration of data are quintessential endeavors in medical discovery, research and development pertaining to all of the areas described above.

As the volume of data and new, diverse data-sources grow exponentially, there is a growing urgency to advance consensus and methods for broad standards and protocols for data quality, sharing, and privacy. This is particularly important with the convergence of clinical, human science and data science that traditionally draw from different thought processes and methodologies.

Human Data Science can play an important guiding role in these efforts by offering multidisciplinary, integrated and intersectoral disciplines and collaborative frameworks. It also demands, and enables, the use of new technologies and methodologies to make better use of data in answering questions both old and new.

As the volume of data and new, diverse data-sources grow exponentially, there is a growing urgency to advance consensus and methods for broad standards and protocols for data quality, sharing, and privacy.

The path forward for a modernized approach to natural history of disease research

The study of the natural history of disease is not a new discipline, but this discipline has taken significant prominence over the last decade due to the growing understanding of the unmet need of defining the etiology and course of many rare diseases.

Moreover, recognizing the many unresolved questions related to COVID-19 as well as the fundamental uncertainties across many major disease areas, such as cancers and neurodegenerative disorders, it is now time to modernize the study of the natural history of disease as a centerpiece in medical research and development.

This will require new thinking and the development of novel research methodologies as natural history of disease studies venture into the uncharted territories of prodromal disease. It will also require new models for learning and collaboration as scientific disciplines that hitherto have had limited collaboration, will need to find a new language and develop a new lexicon for describing disease in its multicomplex nature. As a foundational element in Human Data Science, this will require integrated, multidisciplinary strategies applying expertise and tools from a variety of areas across clinical science, human health, social and behavioral science, and powered by advanced data science.

We will require new thinking and the development of novel research methodologies as natural history of disease studies venture into the uncharted territories of prodromal disease.

REFERENCES

[1] https://www.cdc.gov/csels/dsepd/ss1978/lesson1/section9.html

[2] https://www.pcori.org/sites/default/files/UMD-NORD-Day-Two-Natural-History-Disease.pdf

IQVIA INSTITUTE FOR HUMAN DATA SCIENCE

CONTACT US

100 IMS Drive, Parsippany, NJ 07054

United States

info@iqviainstitute.org

iqviainstitute.org

©2021. All rights reserved. IQVIA® is a registered trademark of IQVIA Inc. in the United States, the European Union, and various other countries. 02.2021.ENT