An intelligent research platform: Moving from data complexity to research excellence

ORION HEALTH
WHITE PAPER //
092023

The challenges and opportunities of healthcare research through data

In healthcare, research is critical to creating a learning health system that routinely incorporates the latest evidence into clinical care. **But, accurate and comprehensive clinical data is imperative to transform breakthroughs into impactful medical solutions.**

The potential of a learning health system is immense—enhancing diagnostics, enabling precision medicine, advancing drug discovery, improving clinical performance, and minimising errors.

The rewards can be huge for healthcare organisations that successfully integrate data-driven research and insights into their clinical and operational processes. Yet, the journey to harness the power of data-driven insights is riddled with challenges.

One enormous challenge data scientists face in conducting data-driven clinical research is research data management – issues accessing clinical data, using time and resources effectively when handling the data, encountering data that is unstructured, low quality, and not uniformly formatted, and at the end of the day, translating all that research data into everyday clinical practice.

Clinical teams often spend valuable time overcoming these complexities, resulting in minor delays and major setbacks, which intensify as healthcare systems accumulate more diverse and voluminous data sources. It's time to navigate this terrain and transform healthcare research.

Unveiling data challenges: Dr. Jade's journey

Embark on Dr. Jade's journey, a Data Science Researcher facing common data hurdles in clinical research. This paper explores how an advanced data platform with intelligent capabilities enhances research efficiency and accelerates adoption.

Meet Dr. Jade, a data scientist collaborating with clinicians and peers to develop a localised readmission risk model tailored to her hospital's context. Her role revolves around pioneering data-driven research methodologies for this project.

Jade starts her research with in-depth conversations with local clinicians to understand the factors contributing to hospital readmission. She strengthens her research by conducting a detailed literature review on the causes of readmissions globally.

When Jade feels she has completed sufficient clinician interviews and literature research, she turns next to clinical data to glean further insights. However, Jade must first gain access to the clinical data she needs to support her research, collect, structure, and process that data, and finally, know what to look for in the data.

Simple, right? Unfortunately, Jade has learned that knowing what you need and being able to do it are two different things.

billion clinical

documents are produced annually

zettabyte (a trillion gigabytes)

of health data generated annually

of an individual's health outcomes are determined

by social factors

The fragmented state of healthcare data

In the United States, about 1.2 billion clinical documents are produced annually. Additional volumes of health-related information are collected from the adoption of wearable tech. **In total, the healthcare system generates approximately a zettabyte (a trillion gigabytes) of data annually, and this amount is doubling every two years.** With such enormous volumes of data available, there is immense potential for enhancing patient value. Yet, it also presents a monumental challenge for professionals like Dr. Jade.

In addition to volume, the ecosystem around healthcare data is complex, with thousands of institutions involved in collecting, transferring, and using patient information. Clinical data can be collected from many sources, such as Electronic Health Records (EHRs), health registries, clinical trials, genetic information, wearables, care management databases, scientific articles, billing, and more.

Lack of access to clinical data

Data scientists confront their first major hurdle: limited access to the essential clinical data vital for their research. Rich clinical data stores exist but are held in many siloed healthcare systems. In today's healthcare landscape, researchers can access no single information repository researchers can access. This is an enormous challenge in the field of clinical research, as in many instances, data scientists and researchers must access data beyond what is hosted in their health system.

Consider Dr. Jade, seeking insights into the local population under study—particularly patients with chronic conditions—to analyse readmission rates. To do this, she must undergo the labor-intensive process of conducting multiple queries to pull broad data sets from various sources, followed by completing time-consuming reports.

Furthermore, Dr. Jade has to meet legal and regulatory requirements when working with health data. In the United States, HIPAA requires an Institutional Review Board (IRB) to approve data use, which adds another step – another layer of complexity and process – to her research project.

Unstructured and low quality data

Hospital records are the building blocks of health data science research. They are rich repositories of patient information collected over many years, often across at least several different patient providers and many interactions with the healthcare system.

Although hospital records are a treasure trove of information for research, they are often so loosely catalogued, cross-referenced, and integrated that the mere assemblage of unstructured data into structured cohorts can take weeks and months.

Non-standardised data across disparate systems

Valuable data can usually be found in large data repositories – often housed on legacy technical platforms. However, it is often a challenge to pull data from these platforms, many of which are still operational and have high transactional volumes.

Creating a comprehensive patient record involves harmonising diverse data types: medications, labs, procedures, and structured and unstructured data—clinical notes, images, diagnostic results, and patient attributes. Dr. Jade's task involves the initial cleansing and standardisation of patient data, a complex yet essential step to enable seamless analysis.

Different types of data, such as social determinants of health

Recent years have seen the rise of social determinants of health (SDoH) awareness—an acknowledgment of non-medical factors like employment and support systems impacting health.

Across many studies authors agree that social determinants of health determine approximately 40% of an individual's health outcomes, morbidity, and mortality.[1]

Dr. Jade's research involves SDoH as part of comprehensive patient records, but the array of data types is staggering. Besides clinical data, Dr. Jade seeks behavioral insights, claims, patient-generated info, genomics, and various "omics" data (e.g., transcriptomics, proteomics). Environmental and exposome data also play a role.

Yet, accommodating them in aggregation remains a challenge, given the limited tool flexibility for new data like SDoH.

1 World Health Organization. (2019, May 30). Social determinants of health. WHO. Retrieved June 28, 2023, from https://www.who.int/health-topics/socialdeterminants-of-health#tab=tab_1

Privacy and security of health data

Health systems are rightly protective of their data, as the privacy and security of health data are paramount in any clinical research.

Data from medical institutions, vital for research, undergoes rigorous checks with federal and local regulations to ensure accuracy and regulatory alignment.

In Dr. Jade's case, sourcing readmission risk rates necessitates IRB-approved paperwork, validating the data's research eligibility.

In some situations, health systems are not properly resourced to facilitate appropriate use of data by researchers. This can sometimes result in overprotection of data and researchers being unable to work efficiently.

Duplication of research efforts

Research data accumulation and reuse requires the ability to import and transform data from various data sources. After all the challenges data scientists go through to maintain data integrity and safety, the data they use in a research project may be thrown out or lost at the end of a research initiative.

Meaningfully curated data used for one project might not be suitable for another, so the process must start all over again, and the effort of the researchers is duplicated.

The solution: A new way to empower data-driven clinical research

Amid the challenges of fragmented healthcare data and data acquisition hurdles, a fresh approach is needed to empower the healthcare landscape.

The health industry needs a solution to ensure the right data ends up in the right hands at the right time.

The next section of our white paper delves into how an advanced data platform, enriched with intelligent capabilities, addresses fragmented healthcare data. It facilitates data scientists and researchers in saving time, resources, and costs while enhancing research efficacy.

Right data, right hands, right time

A modern intelligent data platform should simplify delivering precise data to designated recipients when needed.

The **”right data”** encompasses information from diverse sources—traditional (like clinical data) and non-traditional (like SDoH)—interlinked seamlessly across care stages.

The **“right hands”** imply secure population or patient-level data transmission from sources to authorised researchers, upholding strict patient privacy.

The **“right time”** signifies frictionless data exchange, enabling seamless assembly and timely queries. This is achieved through interconnecting datasets, revealing their intersections for meaningful insights.

Introducing a cutting-edge intelligent data platform

A modern data platform should enable data scientists and researchers to use their data and intelligence assets more effectively. Designed for scalability and openness, this platform should harness modern technology to aggregate diverse health data. It seamlessly captures and manages traditional and emerging data types with embedded intelligent features. This fusion ensures the swift translation of research insights into operational applications.

A single data access point for research

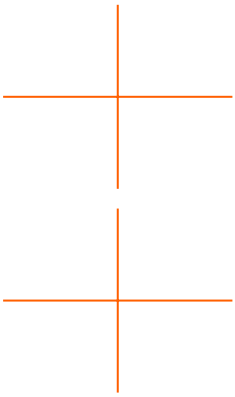
A modern intelligent data platform must act as a customisable data processor, offering pre-built and flexible domains (such as Party, Provenance, Findings, and Business Interaction). These domains facilitate the organised storage of processed data in curated forms.

Furthermore, the platform should seamlessly provide unified SQL query services and APIs. These provisions simplify data access and integration, regardless of data location, whether for research or third-party applications, regardless of data location.

Data scientists engaged in profiling tasks should easily wield data tools such as JupyterHub, compatible with modern machine learning frameworks like Tensorflow. The platform's query services enable the amalgamation of raw, processed, and ad hoc datasets—critical for data profiling, model development, and training.

For instance, Dr. Jade benefits from a centralised data access point via a unified SQL query service. This eliminates the challenge of fragmented datasets, enabling her to securely explore structured, unstructured, internal, and external data sources. The platform removes the need for manual data aggregation.

This modern platform also handles security controls and access permissions cohesively. Unlike the prevalent complex permission management in healthcare systems, this platform centralises access, offering clarity and simplification, including data use permissions and consent, essential for robust auditing.



Streamlined data diversity and accessibility

In a modern data platform, versatility is key. It seamlessly accommodates a myriad of data types from diverse healthcare sources.

For clinical data experts, options abound—fully replicating data or just essential segments, facilitating efficient search, analysis, and reporting. A centralised index simplifies this process. Accessible data resides in a data lake or is directly retrieved from healthcare providers.

This agility extends to new data types—unlike the cumbersome process today. A contemporary data lake is a hub for dynamic data streams (like HL7 results) and static references, supporting collaborative data sets for researchers. Dr. Jade's research exemplifies this; utilising the data lake, she effortlessly engages in data discovery, research, and training, fostering efficiency and synergy while avoiding duplication of effort.

Streamlined model and data management

The intelligent capabilities in this modern data platform, such as machine learning, will provide a central platform where algorithms and trained models can be stored, monitored, and run. This makes it easier for researchers and data science practitioners to manage the life cycle of their algorithms and trained models – both for machine learning and deep learning. In comparison, old systems did not allow for machine learning as a requirement.

Once uploaded and deployed, algorithms and models swing into action, processing incoming data from relevant domains or seamlessly integrating with third-party applications through a REST API interface.

Insights gained from inferences find their place within the platform's domain or extend into external applications, seamlessly merging into clinical or operational workflows without delay.

The platform needs to be uniquely designed to future-proof the evolving state of healthcare data, from traditional data, such as results and encounters, to emerging data, such as social determinants of health and genomics.

This fast-tracks the adoption of the latest insight, as you can rapidly promote validated models to operational use using machine learning. Algorithms and trained models are easy to deploy, integrate, and monitor, allowing model insights to be embedded with workflows in real-time.

Seamless transition from research to practice, empowered by data

Today, the separation between research and practice makes it challenging to graduate from research to practice easily. The modern data platform will empower data scientists and researchers to quickly innovate and deliver timely, intelligent insights. By constructing data science assets that forecast outcomes, automate processes, and prioritise tasks, the platform facilitates the application of models, even those developed abroad, for validation and precise customization.

For example, the nzRISK model, tailored to estimate surgical mortality risk, emerged after evaluating international models on an extensive New Zealand dataset.

Research findings show that effective readmission interventions that are poorly targeted can still reduce hospital readmission rates by over 28%

Example: Predicting inpatient readmission risk[2]

Let's review a real-life example of how the capabilities of a modern-day data platform can help make clinical research more effective.

Predicting hospital readmission risk is important to help identify which patients would benefit most from care transition interventions, as well as to risk-standardise readmission rates for hospital benchmarking.

Research findings show that effective readmission interventions that are poorly targeted can still reduce hospital readmission rates by over 28%.

So, imagine the potential impact of using machine learning techniques to target better and create more personalised interventions for those patients at greatest risk.

Recent studies have shown that readmission risk models must be tailored to local populations and healthcare settings. Subsequently, this real-life research project aimed to validate existing candidate models and tailor a readmission risk model to a local district health board's population and hospital models.

Using a modern data science environment, data from many different hospital and patient domains was analysed. This analysis focused on readmission risk factors available to clinicians at the end of the first day of admission.

These risk factors were:

- Demographics
- Diagnoses
- Procedures
- Medications
- Comorbidities
- Recent admissions and health system interactions
- Test results

For this research project, the data under analysis was first cleaned and validated with clinicians and data experts, then used iteratively develop a predictive machine-learning model for 30-day readmission risk for patients.

To validate the model performance, it was run silently in a cloud environment for months. The model scored new patient data daily, and results were stored for evaluation and tracking through internal analytics dashboards.

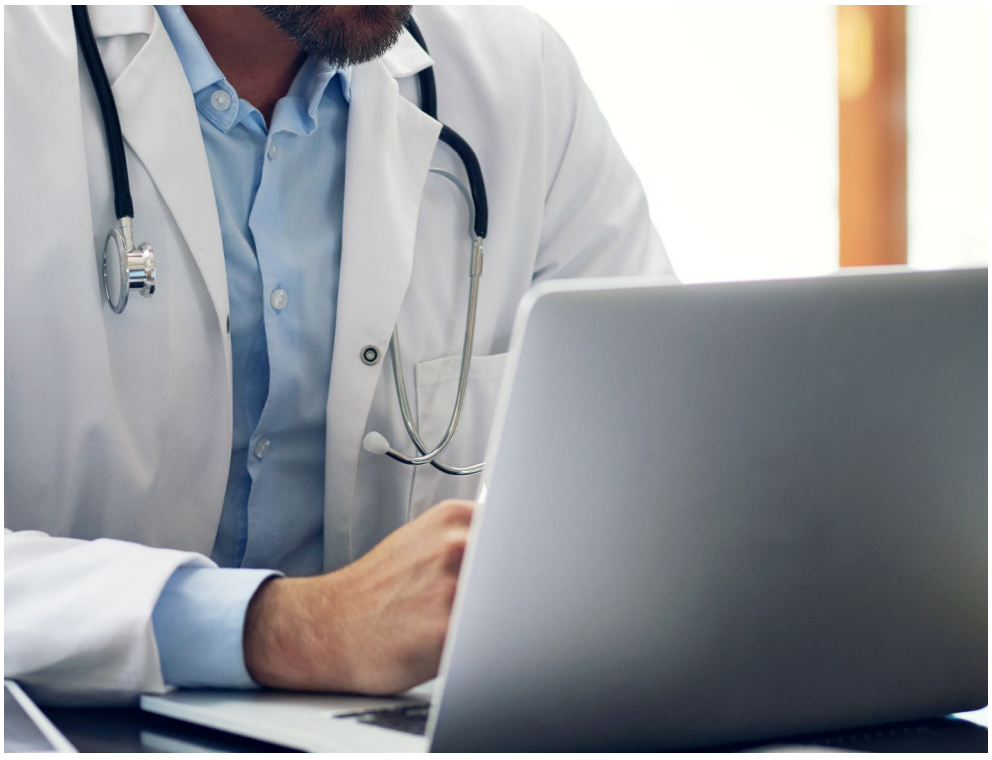
The project aims to seamlessly integrate patient risk scores into clinical workflows. This involves automating risk scores via electronic health record (EHR) data, empowering clinicians with clinically relevant and user-friendly insights.

2 Orion Health and Waitemata District Health Board have developed and deployed a predictive model for readmission risk through a Precision Driven Health-funded research project.

Conclusion

Leveraging vast health data sources holds immense promise for enhancing clinical research, leading to better patient outcomes and streamlined health systems. However, our journey to a learning healthcare system is not without challenges.

This white paper has highlighted some of the challenges and opportunities for machine learning in the field of clinical research. A modern data platform, with built-in intelligence capabilities can transform research effectiveness and support faster and safer adoption.



References

Just, E. (2019, January 17). The 3 Challenges of Translational and Clinical Research Data Management and a Strategy to Succeed. Health Catalyst. https://www.healthcatalyst.com/addressing-challenges-clinical-research-data-management#:%7E:text=However%2C%20researchers%20are%20facing%20problems,data%20into%20everyday%20clinical%20practice.

INTRO

Editor. (2020, February 27). 7 Ways Data Science Is Reshaping Healthcare. AltexSoft. https://www.altexsoft.com/blog/datascience/7-ways-data-science-is-reshaping-healthcare/

Crapo, J. (2019, July 29). The Practical Use of the Healthcare Analytics Adoption Model. Health Catalyst. https://www.healthcatalyst.com/practical-use-healthcare-analytics-adoption-model

Mukherjee, M., Cresswell, K., & Sheikh, A. (2021). Identifying strategies to overcome roadblocks to utilising near real-time healthcare and administrative data to create a Scotland-wide learning health system. Health Informatics Journal, 27(1), 146045822097757. https://doi.org/10.1177/1460458220977579

DATA QUALITY

Larsen, T. (2021, February 25). How to Run Analytics for More Actionable, Timely Insights: A Healthcare Data Quality Framework. Health Catalyst. https://www.healthcatalyst.com/insights/healthcare-data-quality-4-level-actionable-framework

May, T. (2020, August 13). The Fragmentation of Health Data - Datavant. Medium. https://medium.com/datavant/the-fragmentation-of-health-data-8fa708109e13

Orion Health supplies the world's #1 health data platform,
protecting over 120 million precious patient records worldwide.

FIND OUT MORE AT // **ORIONHEALTH.COM**

ORION HEALTH