HiLabs™

APM Case Study 2023

Abhishek Kumar
IIT Roorkee
20119002
abhishek_k@me.iitr.ac.in

Problem Statement

The **field of pharmaceuticals** faces a critical challenge in leveraging **real-world evidence** for data-driven decision making. With **vast and diverse sources** of data, ranging from clinical records to genomic information, ensuring **data quality**, **conversion**, and **curation** has become a daunting task. Errors and **inconsistencies in these datasets** hinder the timely and accurate analysis of patient outcomes, treatment effectiveness, and disease progression. You need to **design an AI** solution that will help solve this problem

Breaking down the Problem

Q. For which industry are we solving the problem ?
Pharmaceuticals.

Q. What is their challenge?
They need real-world data (RWD).

Q. Why do they need data?
They need RWD to draw insights into treatment working in real-life. scenarios i.e, Real-world evidence (RWE).

Q. Why is it a challenge?
1. Presence of abundant data from diverse sources.
2. Inconsistent and unstructured nature of data.
3. Quality issues, i.e., Inaccurate, incomplete and unreliable.
4. Curation issues, i.e., organizing and maintaining.

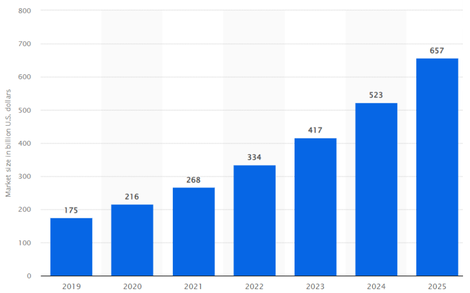
Q. What is the Impact of error?
1. Hinders ability to conduct timely analysis.
2. Gives type-II error in their analysis.
3. Disease progression becomes difficult & unreliable.

Q. Why do we need AI?
1. To improve this data quality, conversion & curation.
2. Complexity & large volume of data.

Q. What all things should also be considered?
1. HIPAA compilance and other medical laws of the US.

Problem Validation

- With the advent of digital health tools, the global digital health market was is estimated to be **~$660 bn+** with an expected **CAGR** of almost **25%** from 2019 to 2025.
- The primary reason of this is the growth of abundant healthcare data which could be utilized for various analytics.
- According to a report by **Mckinsey**, Implementing advanced RWE analytics across the entire value chain for in-market and pipeline products could yield over **$300 mn+** annually for a top-20 pharmaceutical company in the next three to five years.



Ref: Statista

Why are we Solving ?

- The results obtained from clinical trial data make it very difficult to generalize findings to larger, more inclusive populations of patients, providers, and healthcare delivery systems.
- A major **reason** for this is the contribution from external factors like genetics, behavior and environment giving us **low variability in our lab trial data**.
- The generation of abundant amount of health data, from use of computers, mobile devices, wearables and other biosensors, allows us to design better and conduct clinical trials and clinical studies in the health care setting to answer questions previously thought as infeasible.
- With the development of new analytical capabilities, we are better able to analyze these data and apply the results of our analysis to medical product development and approval.
- This however is only possible if we have accurate, complete, reliable and curated **real-world data**.

Understanding RWE

- Real World Evidence (RWE) is the clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD.

Things required to generate RWE

- High-quality RWD source
- Validated approach
- Data analytics capabilities

Process to generate RWE

- The process of creating RWE involves:
  (i) Defining a study protocol to answer relevant clinical questions.
  (ii) Identifying data elements from RWD sources.
  (iii) Establishing data capture protocols within existing RWD sources.
- It is then generated by defining and calculating clinically relevant outcomes and measures, controlling for variability in data quality and confounding patient factors.

Advantages of RWE

- Real-World Evidence (RWE) offers insights into **real-life patient variability**, treatment, and outcomes. i.e., it informs unmet needs, research, and trial design in clinical R&D.
- It **supports healthcare decisions**, safety, effectiveness, and innovation in the industry.
- In sales and marketing, it aids in targeting and healthcare professional (HCP) decision insights thus medical affairs benefit in **pharmacovigilance and product differentiation**.
- RWE helps to understand how patient characteristics impact the health outcomes.
- It can **predict disease progression**, therapy responses, and adverse event risks.
- RWE enhances R&D efficiency and **speeds up time to market**.
- This approach provides **deeper insights** into patient sub-segments, cost analysis, and event prediction, making it a powerful tool for decision-making in healthcare.

Sources of RWD

Clinical

Electronic Health records (EHRs), Labs, Imaging, Genomic, Proteomic, Metabolomic, patient-reported outcomes

Social

Employment, Family, Household & Social networks

Financial

Credit-card spending, Income & Purchases

Behavioral

Diet, Lifestyle & Physical activity

Demographic

Age, Education, Environmental factors, Income & Geographic location

Administrative

Insurance claims & Employment records

Attitudinal

Patient experience & Sentiments

How RWD helps?

- It Increases **patient centricity**, accelerates the pace of scientific innovation, addresses rising development costs, and intensifyies their focus on value.

Challenges with RWD

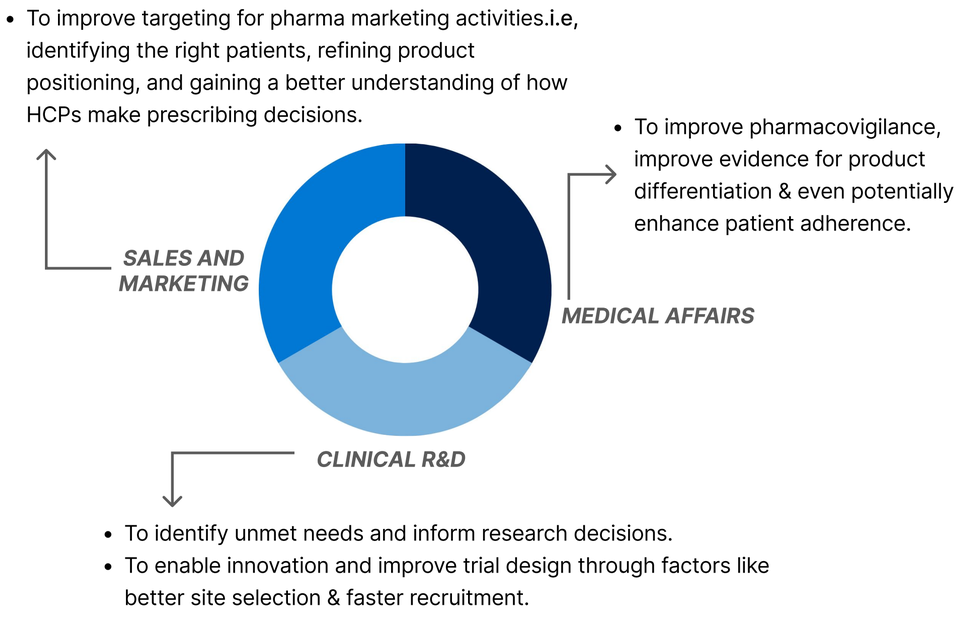
- Data quality, biases, and interpretability are key challenges for the credibility and effectiveness of methods in healthcare.
- Healthcare data sets are typically designed for administrative and billing purposes, leading to incomplete data with inherent biases, and comprehensive data quality assessment is often challenging and costly due to limited visibility into Real-World Data (RWD) until after purchase.
- Biased data can distort results, impacting medical decisions and research.

Stakeholders Involved

- Research & Development (R&D):- To drive drug development and ensure safety and efficacy.
- Healthcare Provider (HCP):-To deliver patient care & making decisions based on clinical evidence.
- Medical Science Liaisons (MSL):- To educate healthcare professionals for new treatments.
- Regulator:- To ensure product safety, approve market entry, and enforce regulations.
- Market Access:- To make healthcare products accessible and affordable.
- Patient:- End-users who are increasingly involved in healthcare decisions.
- Payer:- To cover healthcare costs & assess cost-effectiveness of treatments.

Target User

The target user for this product is a healthcare professional or clinical researcher involved in clinical trials and healthcare studies, responsible for data-driven decisions.



User Research

Persona

Mike
34, Analyst
XYZ Pharmaceuticals

Background

Mike is an analyst at XYZ Pharmaceuticals, he understands the impact that RWE can make however due to the observational, unstructured and high-frequency nature of RWD, he faces difficulty.

He and his team employ traditional methods like Pragmatic clinical trials and Target trial emulation to achieve causal inferences. However, these statistical models often struggle with the large volume of uneven data.

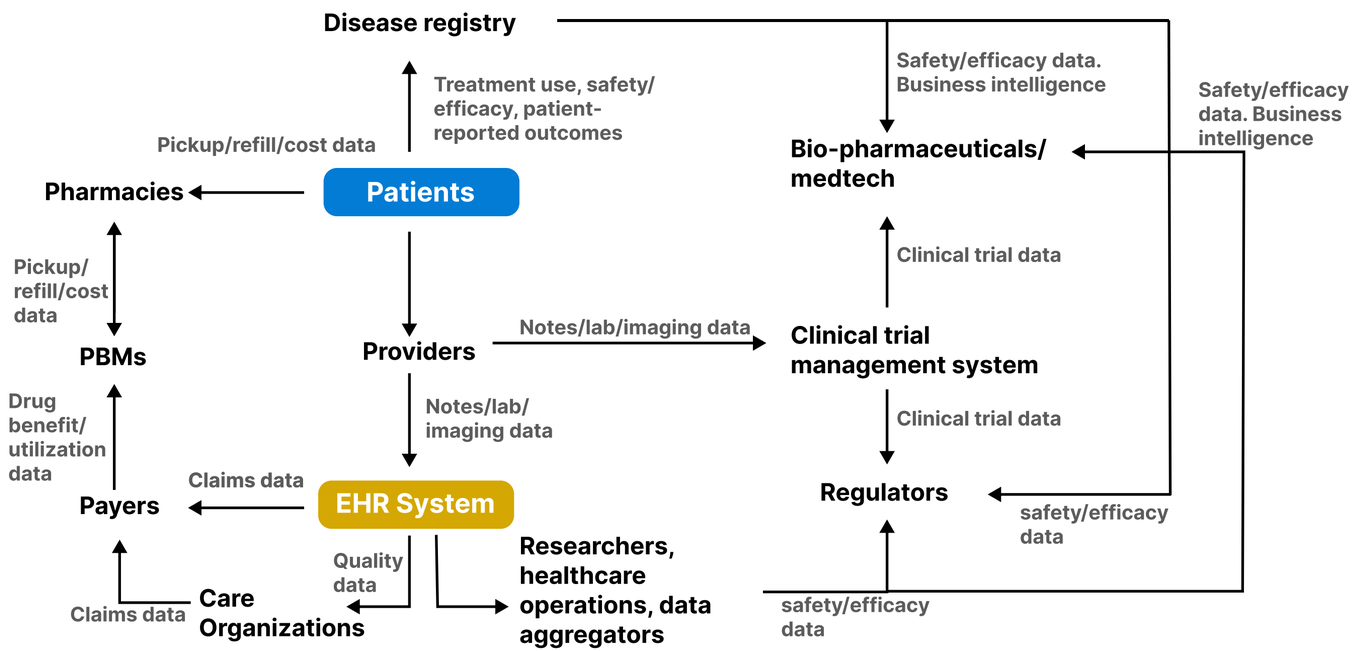
Goals

- To use real-world evidence (RWE) effectively to inform pharmaceutical research and decision-making.
- To achieve more accurate causal inferences, enabling the identification of cause-and-effect relationships in healthcare data.
- To find solutions that can manage and analyze the vast volume of unstructured, and high-frequency RWD effectively.

Pain Points

- Handling high-frequency RWD is challenging, making it difficult to extract valuable insights.
- Traditional statistical models like Pragmatic clinical trials and Target trial emulation struggle to handle the sheer volume of data, limiting their effectiveness.
- Large volume of real-world data poses bottlenecks in data analysis, hindering timely and informed decision-making in pharmaceutical research.

Understanding data generation



For example

EHR Data

Claims Data

When linked and validated with clinical trial findings, post pre-processing,opportunities to learn patterns, make new discoveries, assist preoperative planning, diagnostics, clinical prognostication, alongside improving predictions arise. This also helps to understand patient behavior, drug-drug interactions etc.

Registry data

Improves accuracy of estimates, provides valuable data for supporting regulatory decision-making and enables identification and sharing of best practices, especially for rare diseases.

Useful models

- **SVM:** To predict model relationships between formulation variables (i.e., processing parameters & drug release profile.)
- **K-means:** Chemical similarity, product optimization, market segmentation
- **CNN:** Image based tasks, identifying potential drug targets.
- **Random forest:** Drug discovery and design. drug-drug interaction/ prediction
- **GAN:** Generation of optimized drug candidates, adverse event prediction.
- **ICA:**Dimentionally reduction technique to identify independent components.

Feature Ideation

Data Parsing and Extraction Engine

- A robust data parsing engine that can handle diverse data sources.
- Data extraction algorithms to efficiently gather structured and unstructured data from these sources.
- Natural Language Processing (NLP) techniques for text data extraction from clinical notes.

Data Cleansing and Preparation

- Data cleansing modules to rectify data inconsistencies, missing values, and outliers.
- Data normalization techniques to standardize data formats and coding systems.
- Data quality checks to ensure high data integrity, addressing issues like duplicate records and data linkage.

Data Warehousing and Integration

- A centralized data warehousing system capable of handling vast volumes of data.
- Data integration mechanisms to combine various data types, creating a unified repository for analysis.
- Data version control to track data updates and maintain data lineage for traceability.

Data Governance and Compliance

- Data governance protocols to maintain data privacy, security, and compliance with regulatory standards such as HIPAA and GDPR.
- Access controls and user permissions to ensure data is accessible only to authorized personnel.

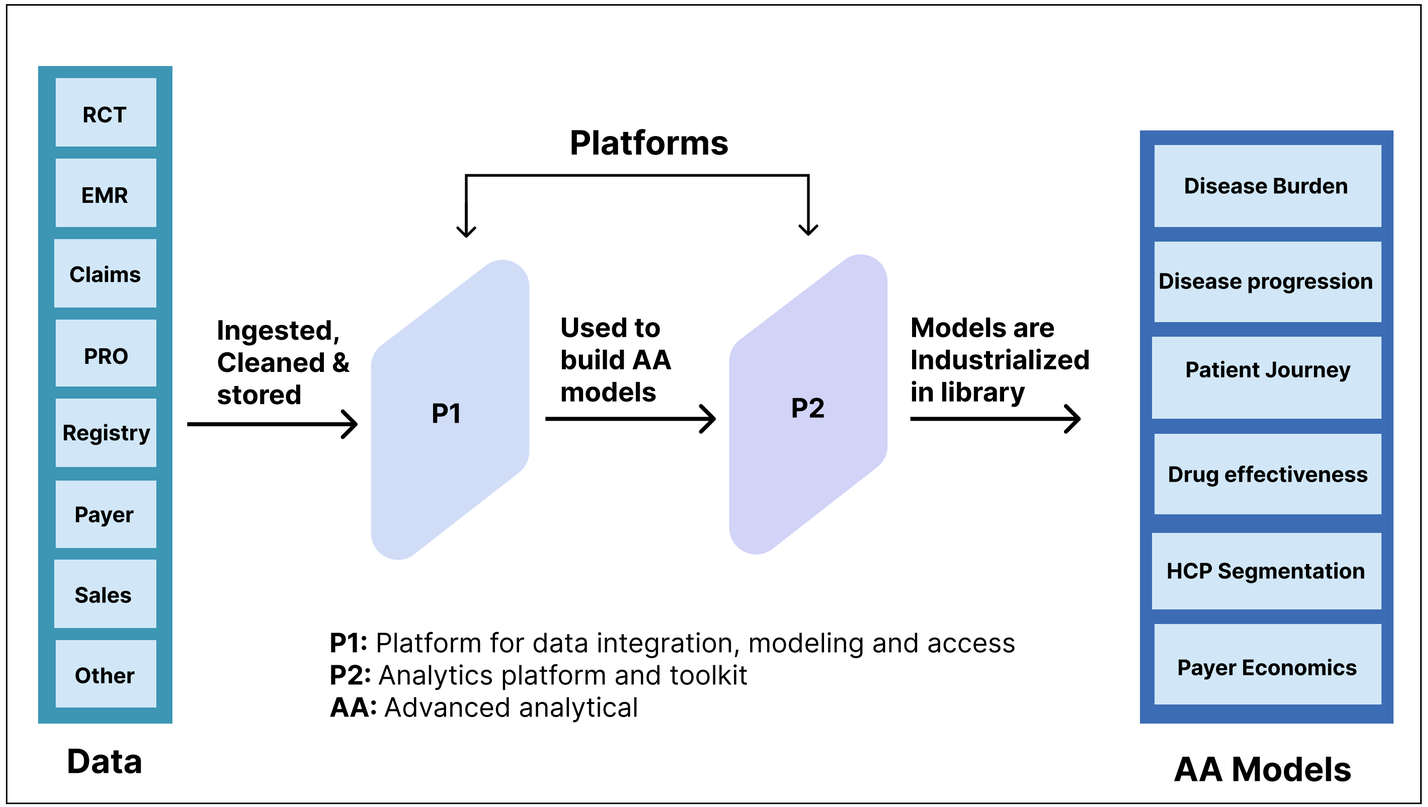
Data Visualization and Reporting

- Data visualization tools to generate dynamic reports, enabling stakeholders to explore and understand evidence derived from RWD.
- User-friendly dashboards for easy access to insights, including disease patterns, patient adherence, and treatment comparisons.

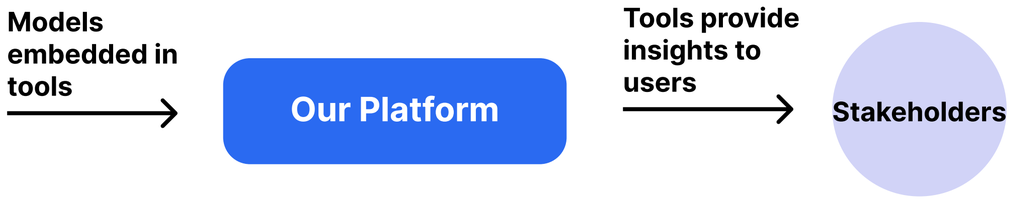
Solution

"We need to create a real-evidence generation engine that would use modern machine learning techniques and harness the power of Real-World Data (RWD). This engine can be offered as a solution in our platform for our stakeholders to get valuable insights in the field of Pharmaceuticals. The engine primarily has two platforms, one for cleaning and storing data and other for performing advanced analysis. The models which get validated are industrialized in the library and are available for our stakeholders as AA models."

Process Overview



Real-evidence generation Engine



Key offerings

- Ability to integrate data from any source without problem.
- Ability to analyze unstructured data allowing to extract meaningful information.
- Ability to exchange HIPAA-Compilant data.
- Biological-pathway insights, revealing new connections for treating medical conditions and improving understanding of disease biology.
- Disease insights, including real-world outcomes, effectiveness comparisons, and in-depth analysis of patient responses, catering to various sub-populations.
- Personalized treatment protocols and insights into factors influencing adherence and treatment changes.
- Features like LCM prioritization, MSL support, clinical decision assistance, beyond-the-pill solutions, and engagement with payers to serve diverse stakeholders.

Wireframes

Product catalog of the MVP Version

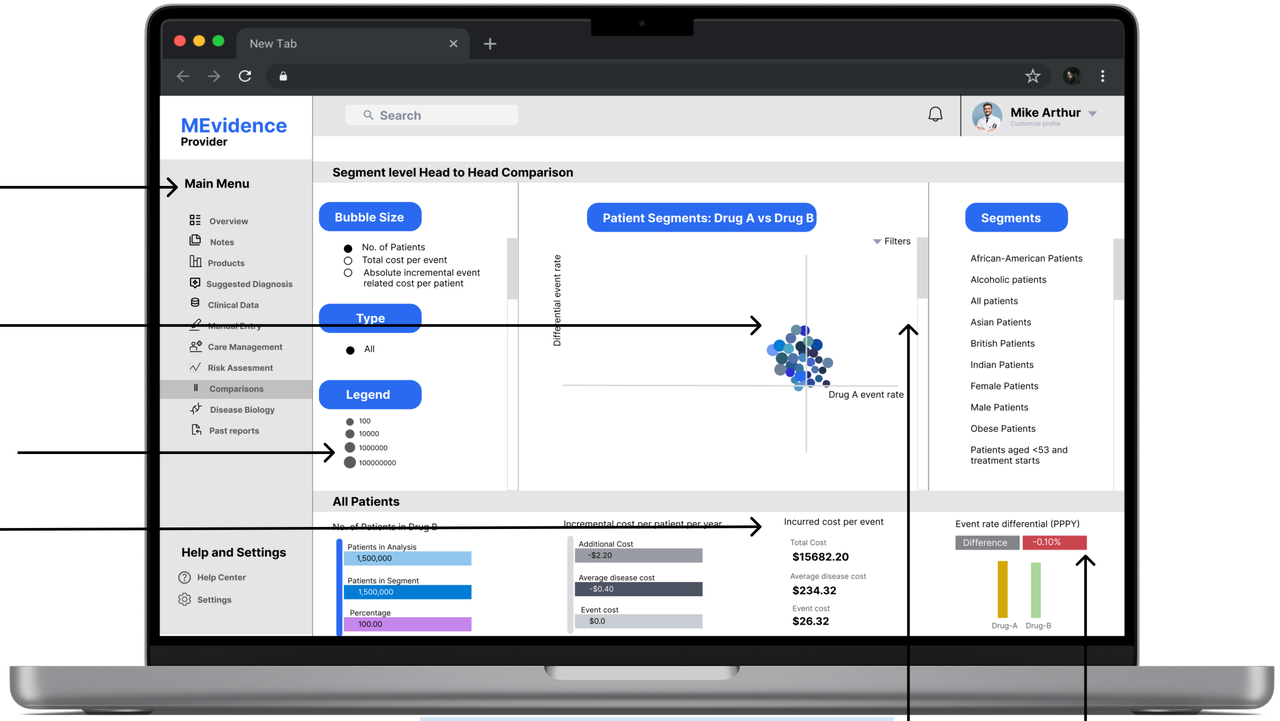
Sub cohorts where drug A outperforms drug B in outcomes but under performs drug A's average across all cohorts

Size of bubble indicates the size of sub-population of the cohort

Cost burden of the event for medical system and the delta cost or savings between compared treatments

Comparison of expected no. of events (eg. falls) with different types of treatments

Sub cohorts where drug A outperforms drug B in outcomes and also ourperforms drug A's average across all cohorts

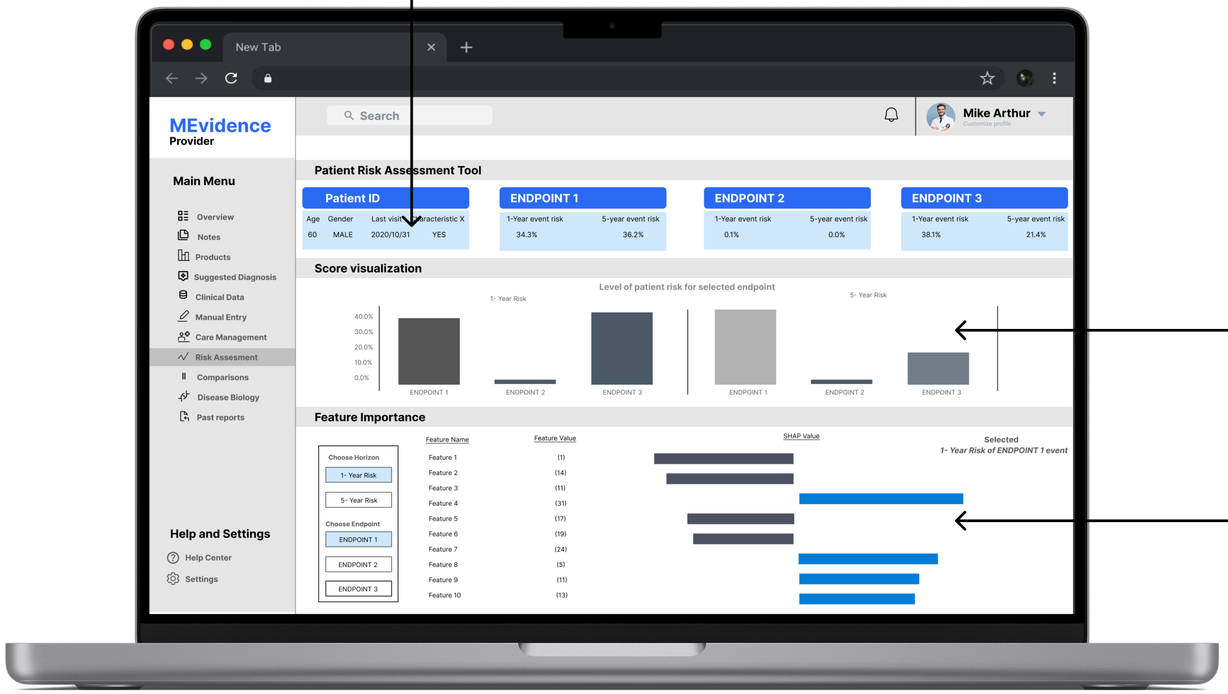


Head-to-Head Comparison

Patient's demographic variables and main comorbidities

Propensity of the patient to experience one of the defined risk factors at endpoints

Review of the main clinical factors contributing/ reducing the estimated risk of experiencing endpoints of interest



Patient risk assessment tool

HiLabs™

Assumptions

- Presence of real-world data on the internet.
- Presence of powerful tools and technologies to format standardize input data and extract valuable insights.
- Cloud-based architecture to store huge amount of data.
- There are support systems for clinical decisions i.e, clinical experts to provide medical inputs and translators to convert business & media requirements into derivatives of technical team.

Pitfalls

- Many healthcare datasets are designed for administrative purposes, not research, leading to incomplete and biased data.
- Comprehensive data appraisal is challenging because data quality and relevance are often only revealed after purchase, making it time-consuming and costly to ensure data reliability.
- Detecting and mitigating biases within the data can be challenging. Biased data can often result in skewed or inaccurate outcomes, impacting medical decisions and research credibility.
- Deep learning techniques can be challenging to interpret as they are often blackbox models, lacking transparency and interpretability.

Future Enhancements

- Predictive modeling and explainable AI techniques can scale up to analyze decision drivers across diverse patient populations and therapies.
- Implementation of Blockchain technology to enhance data security & integrity.
- To ensure that compliance is not breached, strong encryption methods (e.g., SSL/TLS) to secure data in transit and at rest i.e, all patient data, whether stored or transmitted, should be encrypted to prevent unauthorized access along with BAAs and HIPAA Consultation experts.
- Expanding sources of real-world data to provide a broader and more comprehensive dataset for analysis.

Success Metrics

North Star

metric for patient health improvement

- (Improved Health Cases / Total Cases) * 100

Growth

metric to track platform expansion

- ((New Users - Churned Users) / Total Users) * 100

Engagement

metric for assessing user interaction

- (User Activity + Interactions + Content Consumption) / Total Users

Data Quality Index

metric to assess quality of RWD

- (Accurate Data Points + Complete Data Points) / Total Data Points