

Patient's Journey using Real World Data and its Advanced Analytics

Kevin Lee, Genpact

ABSTRACT

Real World Data (RWD) is data collected outside of clinical trial study, and Real-World Evidence (RWE) could be achieved through the insight from RWD. RWD sources come from EMR, health insurance claims, genomic data, and IoT from apps and wearables. RWD anonymized patient data has revolutionized how companies view patient data since it captures longitudinal pharmacy prescription, medical claims, and diagnosis.

The paper is written for those who want to understand how RWD patient data are collected and how they could be analyzed to support pharmaceutical companies. Mainly, RWD patient data could support patient analytics, commercial analytics, and payer analytics such as source of business, switch of prescription, payment method, market analysis, promotional activities, drug launch and forecasting. The paper also discusses the technology that data scientists use for RWD such as Data Warehouse, Data Visualization, Opensource Programming, Cloud Computing, GitHub, and Machine Learning.

What is Real World Data (RWD)

RWD is the data related to patient health and the health care that are delivered to patients. It can come from various resources like

- Electronic Medical Records (EMR/EHR)
- Claim Activities
- Billing Activities
- Product and Disease Registry

Why it is important

RWD could answer some of the questions that cannot be found in Clinical Trial Data, and it helps the life science industry to understand about the industry trends and to make a decision outside Clinical Trial. It also helps the drug companies to define patients' journey, understand drug market status/shift, identify unmet medical needs, and provide the values to the patients. Basically, RWD set up Real-World Evidence (RWE) for the life science industry, and more importantly for the patients.

Real World Patient's Data

Real World Patients' Data is an anonymized patient dataset that captures longitudinal diagnosis, procedures, and treatment history of patients. It also integrates Patients' demographic data, Patients' medical history (e.g., prescriptions, diagnosis, procedures), Payer channel and insurance data, and Physician's information. It could provide one of the most comprehensive Patient Data which could provide history and insight of the patients' medical journey. It could contain 2 to 5 years of patients' activities from his diagnoses, procedures, and treatment history.

Figure 1 shows the journey of patients starting from diagnosis, procedures, and prescriptions.

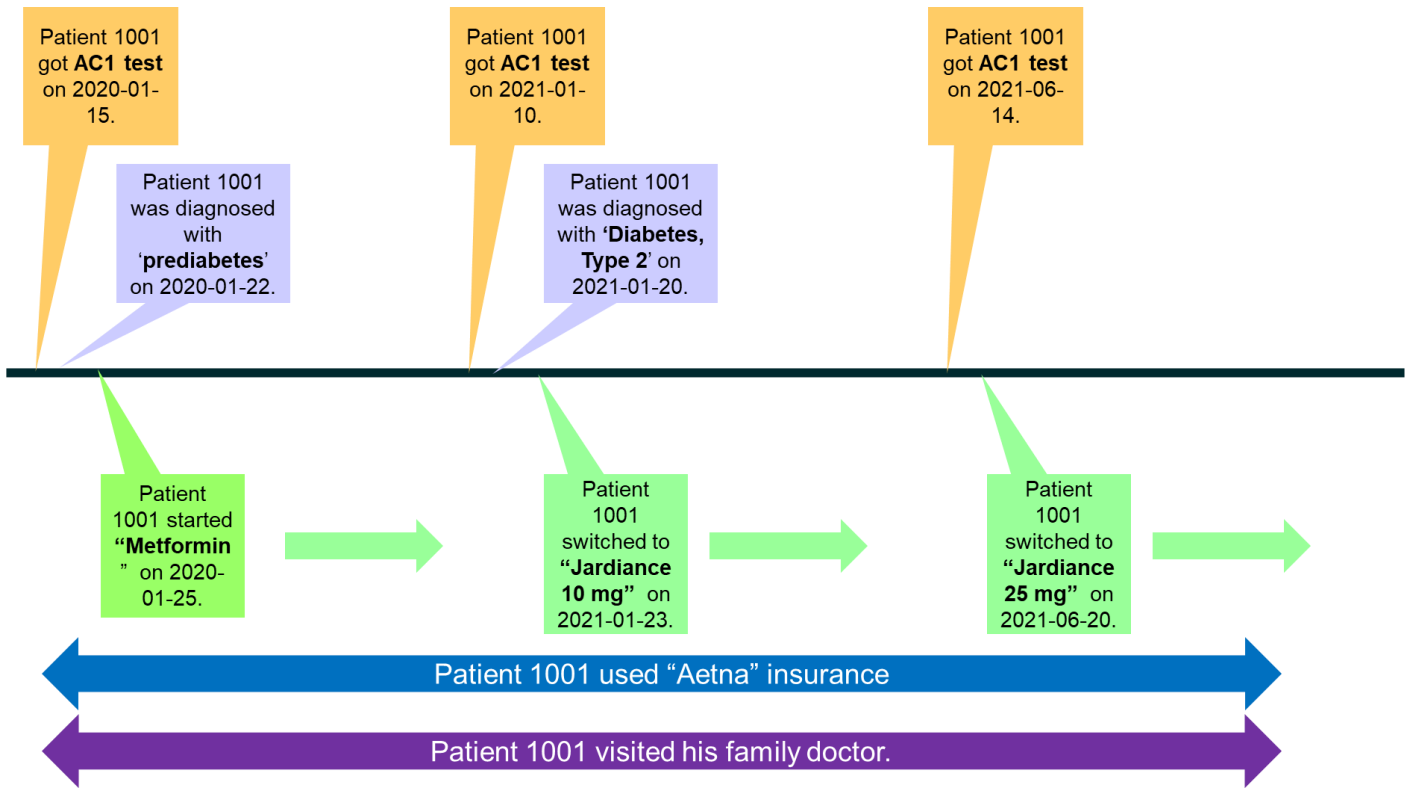


Figure 1. Patient Journey in RWD

Patients' journey in Figure 1 could be saved in different RWD datasets such as patient demographic dataset, patient diagnosis dataset, patient procedure dataset, patient Rx (prescription) dataset, Payer dataset, and healthcare provider datasets. They are all linked with key variables like patient id, payer id, and provider id like ones in Figure 2.

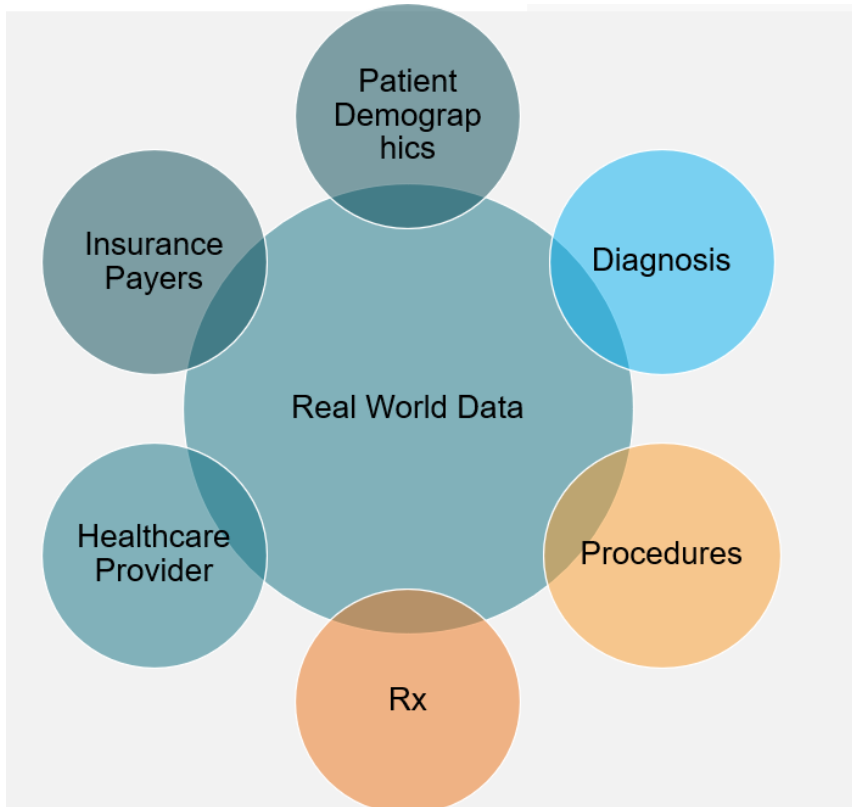


Figure 2. Dynamic contents of Patient RWD

How RWD Data could be used

RWD patients' data could answer the following questions and more.

- Product analysis
- Patient Journey / Activity
- Insurance Payer Distribution
- Prescription prices and Patient's paid prices
- Source of Business (Medicare, Medicaid, Commercial)
- Doctor's activities
- Patient's recruitment for Clinical Trial base on Diagnosis & Procedures

Technology that RWE Patient's Data could be utilized in

The typical RWE patient's data contains 2 to 5 years of patients' medical history data such as prescription, diagnosis and procedure. RWE patients' datasets could contain more than 900 million claims and 90 million patients for each therapeutic area (e.g., Diabetes, Respiratory etc.), which could be considered as Big Data. The analysis of Big Data requires a heavy computing environment. In order to process Big Data faster and more efficiently, it could be recommended to use Cloud computing environment. When using RWD, the data processing could require a couple of steps. For example, data preparation (e.g., sub-setting, filtering) could be done in Database Server (e.g., SQL workbench in AWS Redshift), and advanced analysis (e.g., feature engineering, prediction, modeling, visualization) could be done in Analytical Server (e.g., R Pro Studio/Jupyter in AWS EC2, SAS Server, Databricks in Azure, Tableau Server).

In order for Data Scientists to access and analyze, RWD data should be stored and governed in data storage infrastructure. RWD data also be updated on regular basis (e.g., monthly or weekly) to find the trends of Patients' medical history. There are a couple of options in RWD data storage – physical files (e.g., csv, SAS transport, parquet), Data Lake, Database or Data Warehouse. Since RWD data is governed, analyzed and processed in Big Data and regularly, Data Warehouse is ideal for RWD data storage.

Since RWD data could represent a comprehensive patient's data, it could be used to predict patient's real-world evidence. Prediction model of RWD data could be built using Machine Learning algorithms in SAS, R or Python. RWD data is structured, clean and Big Data, so it could be ideal data to build prediction models. Prediction models of RWD data could predict the future treatment activities, days in therapy, therapy switch, patients' patterns on prescription, diagnosis and procedures, and markets' trends.

RWD data contains years of patient's history data which are integrated with prescriptions, diagnosis, procedures, products, payers, and physicians. It could provide a historical performance such as patient number of certain drugs/ diagnosis / procedures by month, patient's switch pattern by month, geographic display of patients and physician, payer's monthly claims and many more. Such analysis could be clearly visualized by the dashboard using data visualization tools (e.g., Tableau, Power BI, R Shiny).

CONCLUSION

Real World Data (RWD) has become a powerful tool in the life science industry, and it provides valuable business insight and understanding in real world evidence. Especially, RWE patients' data could provide a rich context on patient's medical journey on his diagnosis, his procedure, his prescription, his interaction with physicians and payers (insurance companies), but in order to utilize its value fully, the right technical infrastructure also need to be available such as Big Data Processing Cloud Computing, Data warehouse, Machine Learning, and Data Visualization. With RWE Patient's Data, a data scientist is able to find many aspects of patient's real-world activities, which could also comprehend clinical findings in clinical trials.

REFERENCES

- <https://www.iqvia.com/solutions/real-world-evidence/real-world-data-and-insights>
- <https://www.iqvia.com/locations/united-states/library/case-studies/specialty-patient-journeys>

CONTACT INFORMATION

Your comments and questions are valued and welcomed. Please contact the author at

Kevin Lee
AVP of AI/Machine Learning Consultant
Genpact
kevin.lee@genpact.com

TRADEMARKS

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.

© indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies