PhUSE 2016

Paper RW01

The Playbook: from Real World Data to Real World Intelligence

Laurie Rose, SAS, Cary, NC United States
Robert Collins, SAS, Cary, NC United States

ABSTRACT

Data integration strategies, coupled with innovation in advanced analytics is enabling greater insights into the application of real world data to improve decision making from genomics research to post market studies. This paper will describe the end-to-end process needed for organizations desiring to incorporate observational patient data, claims data, registries, and other sources of data for evidence-based decisions. It will include the preparation of numerous types of data and explore a variety of analysis and visualization methods relevant to therapies and their outcomes across broad populations. We will explore case studies where biopharmaceutical companies are overcoming challenges with big real world data, creating simple to complex patient cohorts, and exploring the attributes of those cohorts. These case studies will explore the insights real world evidence brings to population health management, clinical trial planning, and evaluating the costs and risks associated with care pathways.

INTRODUCTION – THE PLAYBOOK

Real world data (RWD) is global, it's BIG, and it offers a way forward for life sciences companies to better understand outcomes related to the therapies they develop and market. The applications for RWD are vast – from early stage research to planning clinical trials to commercializing the approved therapies. Many organizations are still grappling with the process of leveraging the array of data sources of RWD and using the evidence from RWD to make decisions.

This year's PhUSE follows the 2016 Summer Olympics, where coaches have prescribed to a "playbook" for their sport to take their teams of superb athletes to the gold medal platform. The authors of this paper believe a similar strategy for succeeding in Real World Evidence (RWE) may also require a playbook of sorts. While RWE is not a new concept to the industry, there are new technologies that enable the processing of massive amounts of data and emerging analytical practices that are being applied. It is still a learning process for many, particularly those more attuned to managing and analyzing randomized clinical trial data for safety and efficacy on limited populations.

This paper takes a playbook approach for RWE. The 'rules of the game' are still being formulated by the regulators, so they won't be discussed in detail. At the time of this writing, the European Medicines Agency has several papers around RWD and RWE initiatives, such as the Safe and Timely Access to Medicines for Patients (STAMP) program[(1)]. The U.S. Food and Drug Administration recently published a guidance for medical devices and the use of real world data but not medicines [(2)]. This playbook will focus on The Players (the data), Skills and Drills (managing and preparing data), and The Plays that consist of sample case studies demonstrating the application of RWD analysis and visualization. Figure 1 provides an overview of a typical RWE Playbook.

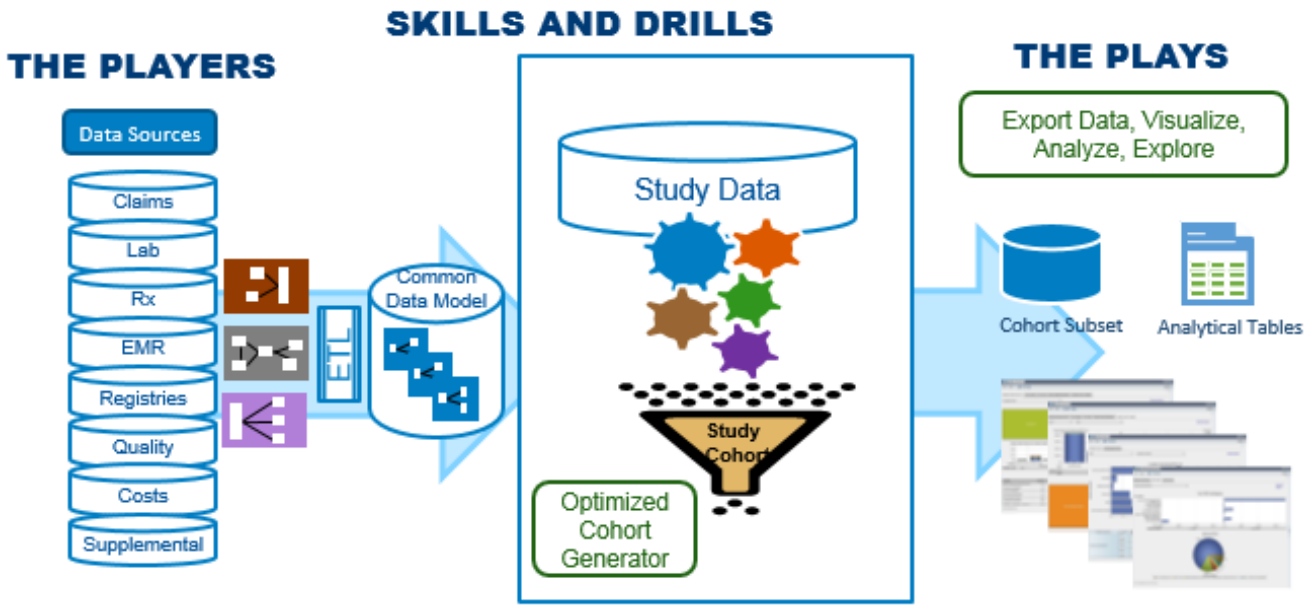

Figure 1 – The Playbook Workflow for Real World Evidence

THE PLAYERS – DATA SOURCES

The availability of Electronic Health Records (EHR) and other digital, health-related data has increased the opportunity for real world evidence research. Data sources are becoming more readily available, including device data (fitness, biometric, etc.), socioeconomic, social media and other health behavior data. Registries can help with the assessment of reimbursement, and Patient Reported Outcomes (PROs) are additional elements to factor into the evidence of efficacy.

Some organizations are adopting a strategy to combine the data sources using a common data standard into a common data model (CDM). The CDM provides a foundation for research methods that "produce meaningfully comparable and reproducible results."[3] OMOP and the OHDSI (http://www.ohdsi.org) organization have made strides in a consortium fashion to help pave the way for researchers with a published open source data model that combines both claims and observational data.

In this playbook, the strategy of a common data model (OMOP or other) is recommended. Certain characteristics of the model are needed for successful observational studies:

- Patient centric
  - The data model revolves around the patient – both clinical interaction data and claims
- Multiple points of entry for analysis
  - Claims
  - Patient interactions
- For efficiency, the model supports both detailed information and analytical summarizations
  - Details support complex cohort generation
  - Analytical data and summarizations can be excluded during cohort selection for efficiency but be available for exploration and analytical modeling
- Manage timelines and refreshes of the data

SKILLS AND DRILLS - PREPARING AND MANAGING RWD

There are a number of considerations when preparing real world data for analysis:

1) Capacity to correlate clinical and financial impacts on patient health

2) Ability to look at index or event driven scenarios as well as population-based cohorts

3) Capability to deliver real-time and historical information
- Current medications, historical medications
- Current vitals, health status, notes
- Longitudinal disease progression for chronic diseases

4) Integration of multiple systems
- Legacy and homegrown systems

- Stand-alone systems
- Facility-specific systems
- Cloud-based 3rd party data systems

5) Differences between real world data (RWD) and randomized clinical trial (RCT) data
- RWD is collected from many sources, including patient reported outcomes
- RCT data follows a strict regimen for completeness and validity

The differences between RWD and RCT data are not minimal. First, there are significant differences in the design and purpose of the raw data from these systems. Real world clinical data is coming from systems designed to capture the daily interactions of patients and providers. Claims data is used for billing and administrative functions. Having poorly specified or missing data leads to a more difficult understanding of the area of study and key related concepts such as diagnosis, treatment and outcomes. As the name implies, RCT is designed to eliminate bias through randomization. RWD will inherently have selection bias not found in RCT data, and could draw into question inferences between exposure and outcome.
If RWD is being used for regulatory decisions based on retrospective analysis of RWD, more rigor needs to be in place to ensure quality and minimal bias.

While still challenging, there is more experience and acceptable methods for imputing missing values in RCT data. For RWD, do records with missing data get thrown out of the game? Are there enough players without those records to finish the game? If not, do those records get sent to rehab to be made whole? These are decisions that need to be made as part of the data quality and data governance strategies that should be a part of every real world data initiative.

THE PLAYS: POPULATION HEALTH MANAGEMENT

A common use case for real world data is to inform and support population health management initiatives intended to improve patient outcomes and reduce healthcare costs. In this case, a payer may be looking to identify patients who would benefit from a care management program.

In the simple example shown below in Figure 2, a care manager is using a data visualization tool to examine a group of patients. Here, the care manager has identified patients with high costs and can see how long it has been since each of those patients have been seen by a primary care physician. The approach is to look at patients who are consuming a lot of healthcare resources. They need to have their care monitored and managed on a regular basis by a primary care provider who can help ensure they have an appropriate care plan that will be followed to better manage their health in the long-term. In this example, the data is filtered to only include members who have not seen a primary care physician in the last 180 days or more.

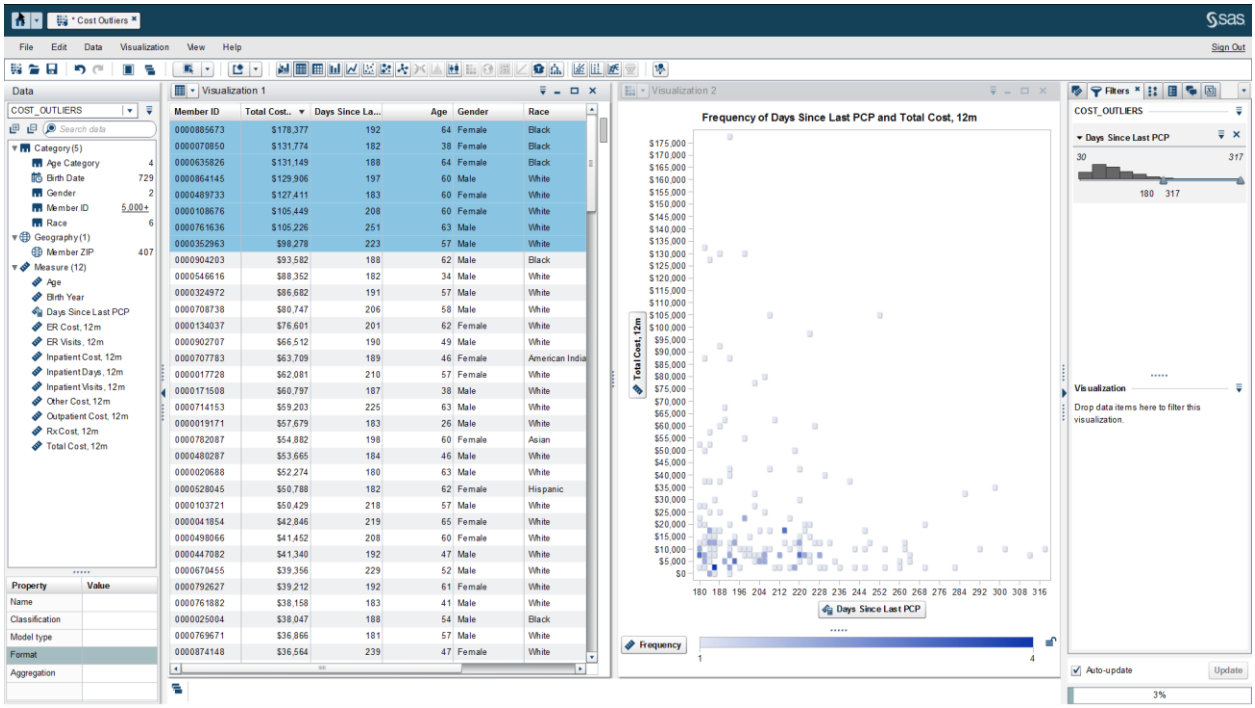

Figure 2: Members with High Cost and Low PCP Utilization

Note that identifying subpopulations of patients who would most likely benefit from a care management program can also be determined by analyzing the same data sources. This process would leverage machine learning or traditional statistical modeling techniques to identify not only patients who are at greatest risk for subsequent events but also their likelihood to respond to various types of interventions such as phone calls from nurses, support groups, emails, text messages or other mechanisms of outreach. With a flexible visualization tool, a user could explore data in an ad hoc fashion to discover unexpected trends and issues while reporting capabilities allow the creation of standardized displays. The reports could be reviewed on an ongoing basis to monitor the status and activity of the entire population or just selected segments, e.g., newly discharged patients, those in high-risk groups, etc.

THE PLAYS: CLINICAL TRIAL PLANNING

Planning a clinical trial requires a large number of complex decisions that are often guided by experience and instinct. Having a platform that incorporates real world evidence provides the opportunity to bring additional rigor and quantitative analysis to the process. In the example presented here, the contract research organization (CRO) needs to identify sites that should be targeted for participating in an upcoming study. One question related to this process is if there is a need to recruit additional sites beyond those with whom they already have an existing relationship and, if so, are there geographic locations where they should focus.

In Figure 3, below, the CRO has applied the anticipated study criteria to longitudinal claims and clinical data to gauge the number of people who may be eligible for participation. These charts show summary characteristics of the population matching the study criteria such as age, sex and race distribution. The researcher must keep in mind that these data reflect only that portion of the population that is included in the available data source. The results being viewed could be biased in terms of age, geographic distribution, income, education level or many other factors that could affect actual study eligibility and recruitment.

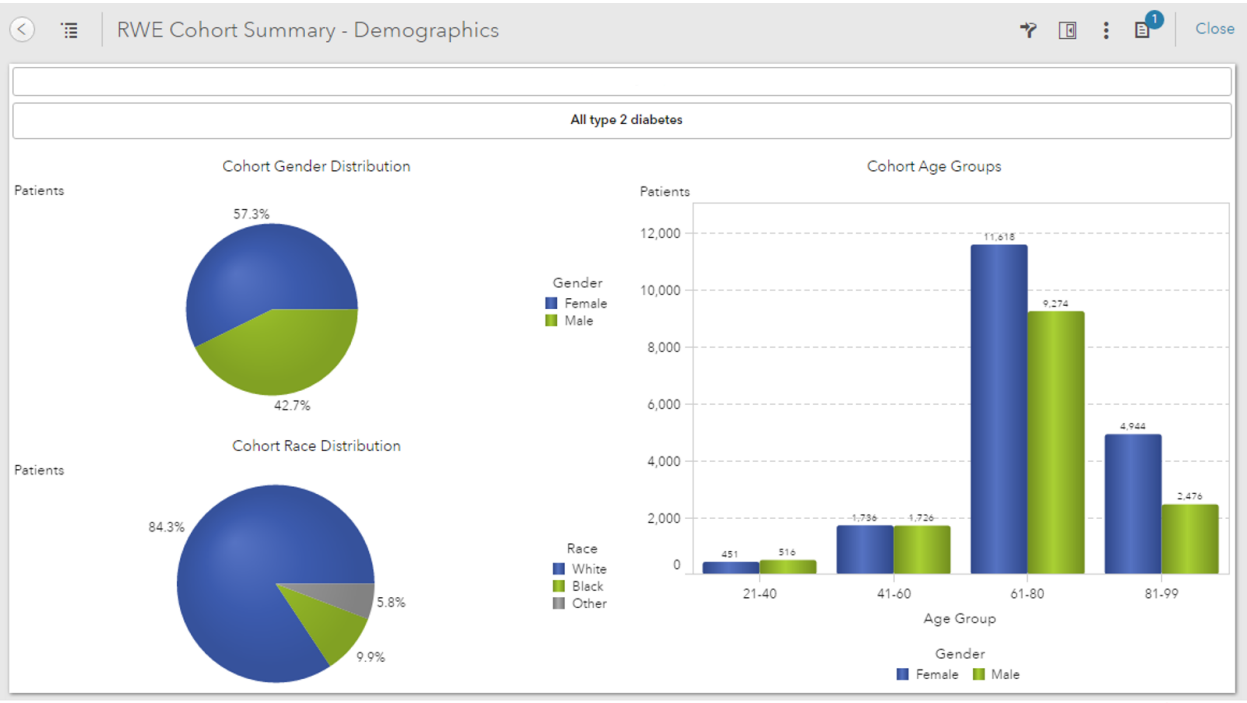

Figure 3: Demographic Summaries of a Cohort of Patients Meeting Anticipated Study Criteria

Figure 4, below, shows the geographic distribution of these potentially eligible subjects as a bubble geographic map. In this case, each bubble represents a US ZIP code where the size indicates the number of potentially eligible subjects in the data and the color indicates if the CRO has an existing relationship with a clinical site in that same geographic area. In this map, large bubbles represent large numbers of potential subjects so nearby sites (indicated by blue bubbles) may be able to recruit from the surrounding areas.

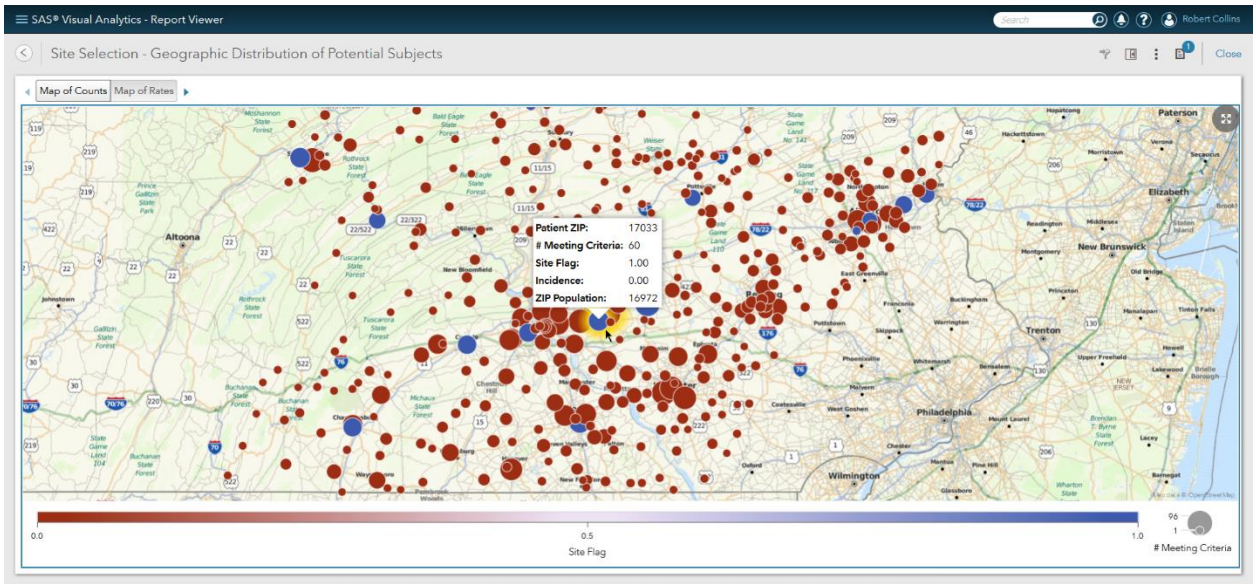

Figure 4: Bubble Map Showing Potential Subject Counts (Size) and Presence of an Existing Site Partner (Yes=Blue; No=Red) for US ZIP Codes

Figure 5, below, is a similar graph to the one above but instead of the bubble size representing the actual numbers of people, the bubble size represents the prevalence of individuals meeting the proposed criteria calculated when merging in census population data. Here, large red bubbles with no nearby blue bubbles indicate areas that warrant investigation to determine if a new site should be recruited in that region.

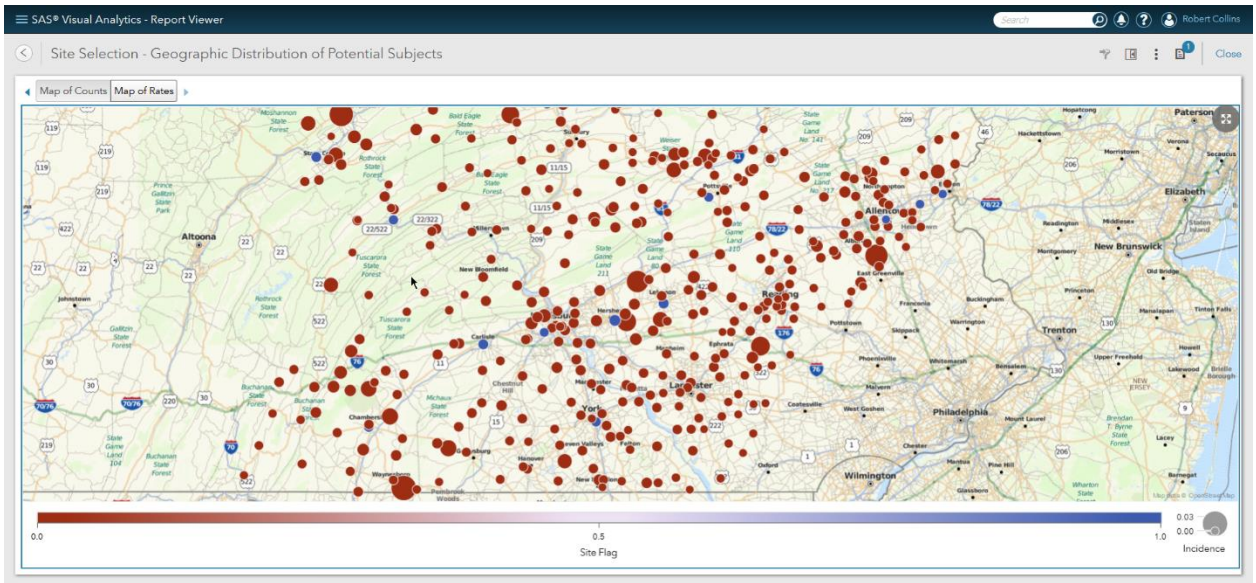

Figure 5: Bubble Map Showing Prevalence of Potential Subjects (Size) and Presence of an Existing Site Partner (Yes=Blue; No=Red) for US ZIP Codes

Taken together, these plots allow study planners to identify any geographic areas where they may want to select or recruit sites and subjects, based either on the number of potential subjects or based on areas with an elevated prevalence of potential subjects.

While the example plots were created using US ZIP codes, other geographic measures or administrative boundaries such as country or region could be employed. In addition, it may be informative to allow the planner to see how many sites are within a specified distance from the location, e.g., color the bubbles by how many sites are within 50 or 100 kilometers, etc.

THE PLAYS: EVALUATING COSTS AND RISKS ASSOCIATED WITH CARE PATHWAYS

Combining a variety of data sources across a large population of patients with a tool that allows rapid insight, analysts can discover and monitor many factors affecting the health of a population and make inferences to the broader populace, as well.

In Figure 6, below, an analyst is looking at patients within an identified therapeutic area see if there are different responses to different therapies. In this case, the analyst is comparing her company's drug in the decision tree on the left with all other drugs for the same condition on the right. This example shows that the factors affecting cost of care are different between the two sets of drugs. For the therapy provided by the analyst's company, a history of coronary artery disease is the first most important factor, whereas a history of congestive heart failure is the most important factor for the other therapies.

In both decision trees, the outcome of interest under consideration is total healthcare cost but it could be other measures such as initial treatment selection, lab values, other indicators of treatment response, indicators of disease progression, event-free survival times, or anything else of interest to the researcher.

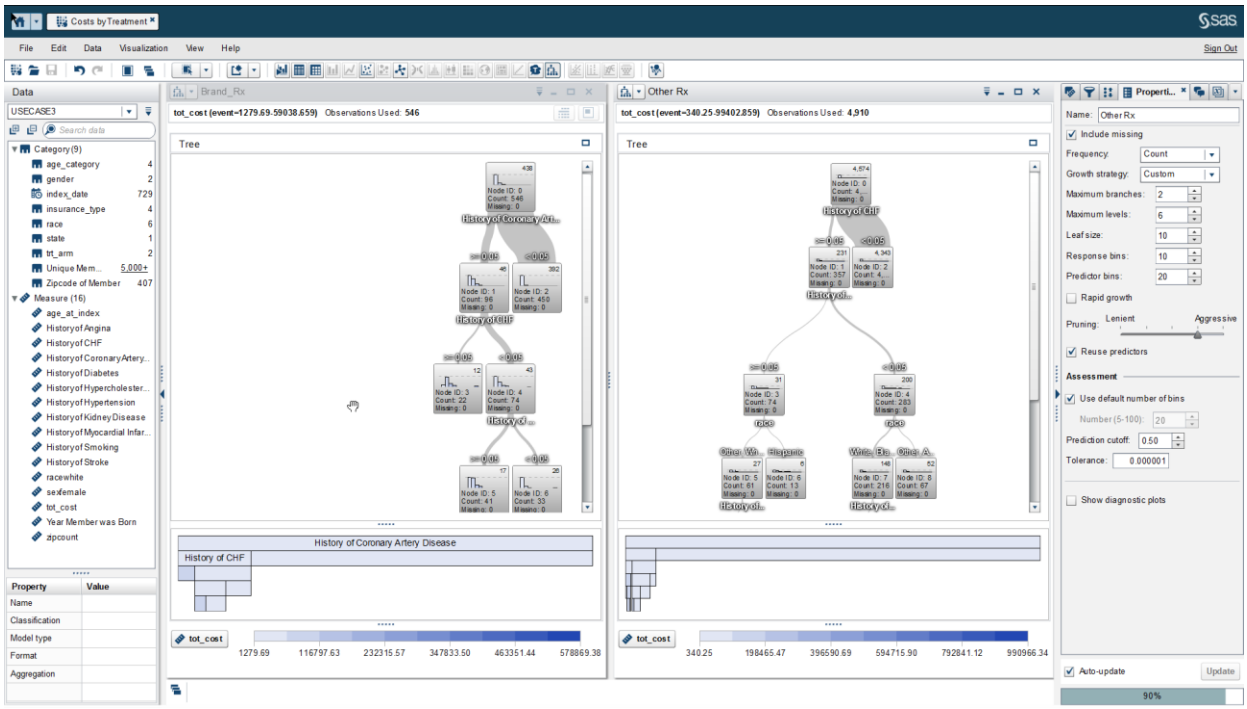

Figure 6: Decision Trees to Assess Factors Affecting Cost for Two Groups of Treatments

CONCLUSION

Real world evidence initiatives are here to stay and will play an increasingly important role affecting all participants in health care. Data science and innovative advanced analytical methods fit well into the world of RWE and add confidence to gaining deeper understanding of disease, diagnoses, treatments and outcomes. "Gold medal results" can come from care decisions and business decisions made by recognizing and reacting to the information gained by exploring and analyzing these data sources.

Advancing decisions in 'the plays', will depend on perfecting the basic skills required to manage and govern the data sources and playing within the rules and guidance provided by standards and regulatory agencies. Just as athletes improve their skills by constantly practicing the basics, the tools and methods for working with RWD will evolve to make it easier and quicker to work with these data sources and improve the quality of the source information. Tight controls around data capture that are expected in RCTs may evade us for a long time in RWD. However, improved systems and analytic methods will allow us to extract increasingly more meaningful information that will benefit patients and the health professionals who care for them.

REFERENCES

1. EMA (European Medicines Agency). March 1, 2016. *Paper for STAMP meeting 10 March 2016: Update on Real World Evidence Data Collection*.
http://ec.europa.eu/health/files/committee/stamp/2016-03_stamp4/4_real_world_evidence_background_paper.pdf

2. U.S. Department of Health and Human Services, Food and Drug Administration, July 27, 2016. *Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices, Draft Guidance for Industry and Food and Drug Administration Staff.*
http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf
3. Observational Health Data Sciences and Informatics (OHSDI). 2014. *OMOP Common Data Model Specifications Version 5.0.1.* http://www.ohdsi.org/web/wiki/doku.php?id=documentation:cdm:single-page

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Laurie Rose, SAS
SAS Campus Dr.
Cary, NC 27513 USA
Work Phone: 919-531-7124
Email:Laurie.Rose@sas.com

Robert Collins, SAS
SAS Campus Dr.
Cary, NC 27513 USA
Work Phone: 919-531-0051
Email:Robert.Collins@sas.com

Web: www.sas.com

Brand and product names are trademarks of their respective companies.