Paper RE05

Is Your Analytics Platform Ready for Handling Real World Data for Regulatory Grade Real World Evidence?

William YuanHao Kuan, SAS Institute, Tokyo, Japan
Olivier Bouchard, SAS Institute, Paris, France

ABSTRACT

Real-world data (RWD) and real-world evidence (RWE) receive increasing attention in the life science industry. RWE has the potential to save on the cost of research and development, and shorten the time for patients to access innovative treatments. In order to use RWE for regulatory decision making, FDA has issued multiple guidances documents to address the usage, benefits, risks, and consideration of using RWE for both safety and effectiveness determinations. An open and trusted research environment can help an organization leverage the value of RWD through talents with different skill sets, ensuring confidence in the quality, reliability, relevance, and transparency of the RWE generation process.

We discuss here the key considerations of building a RWE analytics platform and long-term strategies in using RWD and generating RWE with analytical capabilities for regulatory submission purposes.

INTRODUCTION

Real-world data (RWD) and real-world evidence (RWE) are hot topics in the healthcare and life science industry. The US FDA defines RWD as data relating to patient health status or the delivery of health care routinely collected from a variety of sources. RWE is the clinical evidence regarding the usage and potential benefits, or risks of a medical product derived from analysis of RWD. In other words, RWD can be defined as briefly as data obtained outside the context of randomized controlled trials (RCTs) generated during routine clinical practice (e.g., diseases/product registries, claims, electronic health record).

Why is RWE a hot topic and why is RWE important to the healthcare and life science industry? RWE has the potential to save on the cost of research and development, and shorten the time to patient by leveraging the rich information RWD can provide outside the context of RCTs. RWE is now emerging as a critical evidence modality that can potentially enable regulatory, reimbursement and provider decisions. It can be used to help design and operationalize clinical trials and support program prioritization. In this paper, the RWE related regulatory landscape in the US, Europe, and Japan, and the preferred key features of a modern analytics platform for analyzing RWD were discussed.

APPLICATION OF RWD/RWE

Unlike data from randomized controlled clinical trials, RWD is data generated during routine clinical practice, for example, electronic health records, insurance claims, and patient registry. RWE is the clinical evidence derived from the analysis of RWD. The HMA/EMA Big Data Task Force in Europe uses the term “big data”, including data from genomics, social media and wearable devices. The Japan Pharmaceuticals and Medical Devices Agency (PMDA) RWD Working Group mentioned utilizing medical information databases as the use of RWD. The medical information databases include electronic medical records, diagnosis procedure combination (DPC) from hospital information systems; claims for medical fees and dispensing fees; and disease registries.

RWE can be applied to support across the life cycle of drug development, including developing strategy, clinical studies, submission to regulatory and post-marketing studies. For example, RWD can be used to check trial feasibility to optimize the clinical trial design and avoid possible protocol amendment, or RWD can be utilized to generate an external control arm to support the regulatory submission. For example, it is difficult to identify a sufficient number of study participants with rare diseases for a RCT; or it could just be unethical, impractical, or infeasible to do so. In some circumstances, using RWD to create a synthetic control arm (i.e., an external control arm) can solve these challenges and provide evidence of safety and effectiveness when comparing with the standard of care.

RWD can also be applied in observational studies to demonstrate a product’s safety profile in a real clinical setting. For example, Pfizer submitted RWE in support of the FDA approval for expanding the indication of Palbociclib (trade name: IBRANCE) to male patients with breast cancer. The RWE was derived from a retrospective outcome analysis using data from an electronic health record database. Although concerns around the quality of evidence existed, the FDA approved the supplemental new drug application in April 2019 and commented, “*Based on limited data from post-marketing reports and electronic health records, the safety profile for men treated with IBRANCE is consistent with the safety profile in women treated with IBRANCE*”.

RWE REGULATORY LANDSCAPE

The approach of using RWD as the evidence for health technology safety evaluation (e.g., post marketing survey) is relatively mature compared with using RWD as the evidence for effectiveness determinations. To have a clearer and consensus approach for regulatory decision making, the FDA has issued multiple guidance since 2017 to address the usage, benefit or risks, and consideration of using RWD and RWE for both safety and effectiveness evidence generation. The list of guidelines can be found in Table 1.

Table 1 List of RWE related guidelines from US FDA

Date Published	Title
August 2017	Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices
July 2018	Use of Electronic Health Records in Clinical Investigations
December 2018	Framework for FDA's Real-World Evidence Program
May 2019	(DRAFT) Submitting Documents Utilizing Real-World Data and Real-World Evidence to FDA for Drugs and Biologics
September 2021	(DRAFT) Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products
October 2021	(DRAFT) Data Standards for Drug and Biological Product Submissions Containing Real-World Data
November 2021	(DRAFT) Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products
December 2021	(DRAFT) Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products
September 2022	Submitting Documents Utilizing Real-World Data and Real-World Evidence to FDA for Drugs and Biologics

As of 2022/10/07

The guidelines cover the scope from basic consideration to submission format and assessing RWD sources. The health technologies covered by the guidelines include drugs, biologics, and medical devices. FDA mentioned the potential use of RWE can be to support adding or modifying an indication or comparative effectiveness or safety information.

In the EU, enabling the use of RWE for regulatory decision-making about medicines in Europe by 2025 is the vision of European regulators. The Data Analytics and Real World Interrogation Network (DARWIN EU) will be the key to delivering this vision. This network will allow to access and analyze healthcare databases from across the EU, and to provide timely and reliable evidence on medicine. For the Japanese PMDA, the current regulation of using RWD is more focused on Post-Marketing Safety Assessment. In 2021, PMDA published a guidance on the use of registries for new drug applications, and the development of more guidelines on general principles and data reliability standards for medical database utilization for new drug applications is expected in the near future.

The Framework for FDA's Real-World Evidence Program was issued in 2018. This is a document to describe why, what and how the FDA consider RWD for regulatory submission purposes. Under the 21st Century Cures Act, FDA's RWE Program must evaluate the potential use of RWD for evaluating product effectiveness to help support approval of new indications for drugs, as well as help to support post-approval study requirements. There are three key considerations when discussing RWD for regulatory purposes as mentioned in the Framework document. First, whether the RWD are fit for use; second, whether the trial or study design used to generate RWE can provide adequate scientific evidence to answer or help answer the regulatory question; and third, whether the study conduct meets FDA regulatory requirements. In other words, it is crucial to prove to the regulatory authorities that the DATA and METHODS being used are reliable and relevant to the regulatory questions, and to show the whole evidence generation PROCESS is reproducible and transparent.

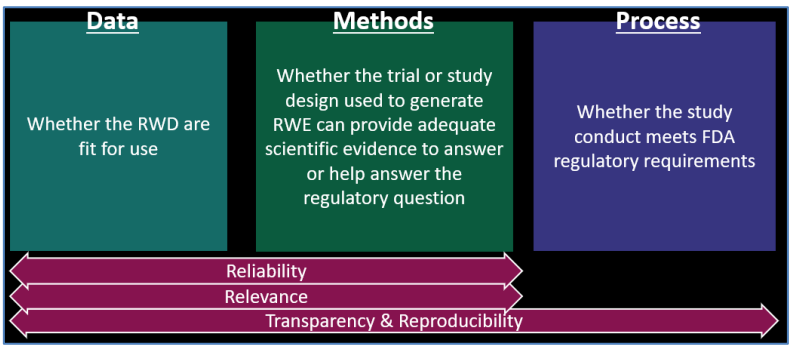

Figure 1. Key Considerations for the Use of RWE in Submission

FEATURES OF A MODERN RWE ANALYTICS PLATFORM – DATA, METHODS, PROCESS

In this section, the key features for a modern RWE analytics platform are discussed from the perspective of DATA, METHODS and PROCESS. A modern analytics platform can help leverage the power of RWD as well as the long-term strategies to make more use of RWD.

DATA – RELIABILITY & RELEVANCE

When assessing the fitness of the RWD, it is important to consider the relevance and reliability of the RWD for answering the regulatory question. First, in order to understand whether the data source we intend to use captures relevant data elements on exposure, outcomes, and covariates, and to understand if the data has an adequate size of population, it is preferable that the analytics platform can **quickly profile large volumes of data through data visualization, and automated advanced profiling**. The user can understand what the data looks like and investigate and discover any data quality issues they might want to address before using the data.

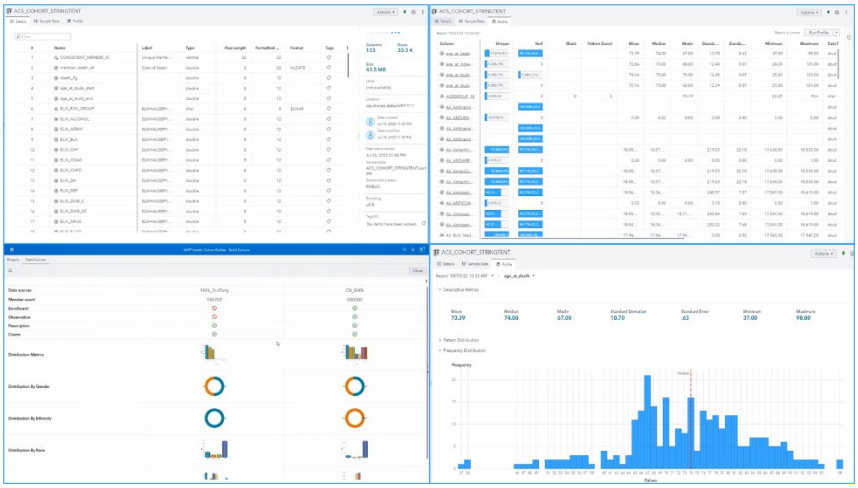

Figure 2. SAS Viya and SAS Health Cohort Builder - Data Profiling

Second, once the user can identify relevant data sources, they also need to ensure the data is reliable. The term "reliability" includes data accuracy, completeness, provenance, and traceability. It is desirable that an analytics platform **has a systematic approach to control data quality**, to provide a clear documentation on how the data was moved, transformed, and enriched, and even can automate the repetitive tasks for data preparation. A transparent process can increase confidence in the resulting data during data curation and data transformation.

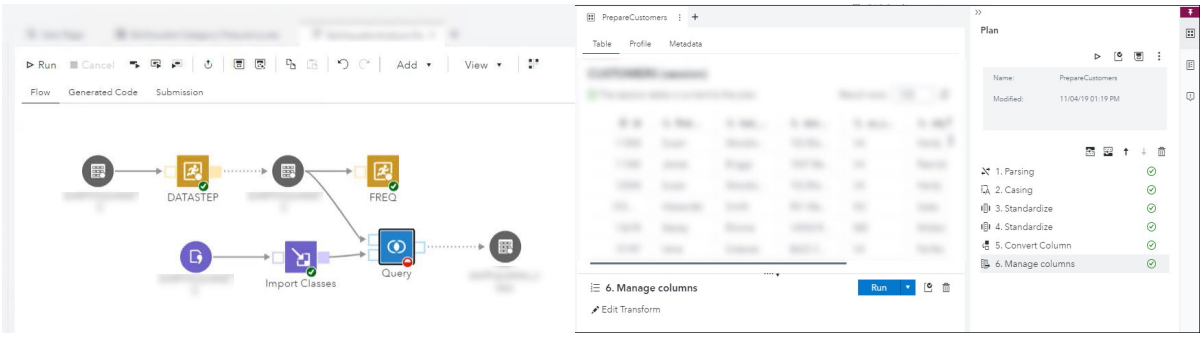

Figure 3. SAS Viya - Flows and Data Studio Plan

Sometimes it is necessary to work with multiple sources of electronic health records, registries, and claims data to answer the regulatory question. It is worth to **consider the common data model or data standards designed** for multiple RWD assets. It is also preferable to **have the ability to efficiently manage common representations** such as coding schemes, vocabularies and terminologies, and to have the ability to reuse the validated study variables within the organization. This allows the research team to gain insights from various RWD assets efficiently and maximize the value of these data. A common data model in place can also serve as the role of supporting data quality check. For example, in the process of mapping raw data into a common data model, the team can handle missing data, data discrepancies, duplicate records and implausible values. The FDA also recognizes the importance of developing data standards to maximize the utility of RWD. However, the users may need to consider the possible loss of information during the mapping process, as not all the info in the raw data can be mapped to a common data model or be reconciled if different data sources are involved.

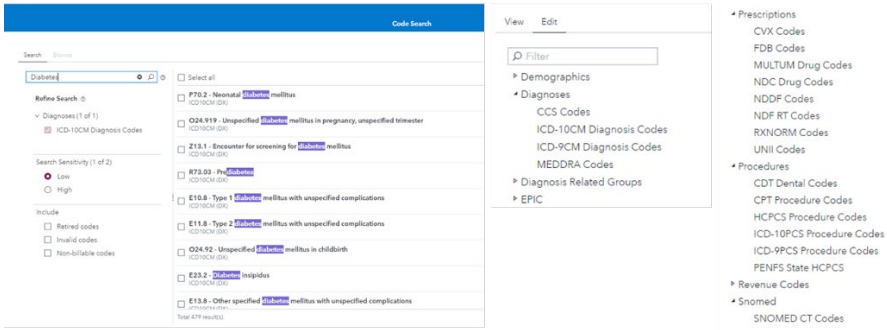

Figure 4. SAS Health Cohort Builder - manage taxonomies and dictionaries

Although electronic medical records, registries, and claims data are commonly used RWD currently, it is foreseeable that the use of more diverse RWD, such as mobile devices, ePRO, eCOA, digital biomarkers and medical images will be more prevalent in the near future to provide deeper insights. The FDA mentioned they will continue exploring strategies to fill the gap of utilizing RWD from biosensors, electronic patient reported outcome instruments, and mobile/wearable technologies. It is recommended that an analytics platform can **consume multiple types of data in large volumes and can connect to a wide variety of environments**. For example, PARQUET and ORC files in Hadoop, or data in the memory, in the database, in the cloud, or real-time data in the streaming processing.

METHODS – RELIABILITY & RELEVANCE

In addition to the fit-for-purpose data, the analytics platform should also be able to help the research team show that the method and study design being used to analyze the data are appropriate and relevant to answer the regulatory questions. Also, the whole process of analysis should be transparent and reproducible.

One of the common approaches to evaluate whether the study design is appropriate is to check if the study design can capture sufficient numbers of representative patients and have adequate statistical precision and power from the data. For example, how do we define the index event? What is the operational definition of the event, outcome, and covariate? These factors could affect the feasibility of the study. When designing the study, it is preferable that the evaluation process, as well as the results from feasibility checks and exploratory analyses are clearly documented.

As the evaluation is a repetitive process, it is desirable that an analytics platform could enable the research team to **quickly identify patient cohorts** with customized inclusion/exclusion criteria and **efficiently manage the feasibility evaluation process in a transparent way**. These features can help demonstrate to the stakeholders that we are not cherry-picking in our analyses. By reducing the turnaround time for feasibility checks and exploratory analyses, it allows clinicians and subject matter experts to have more time to discuss and optimize the study design.

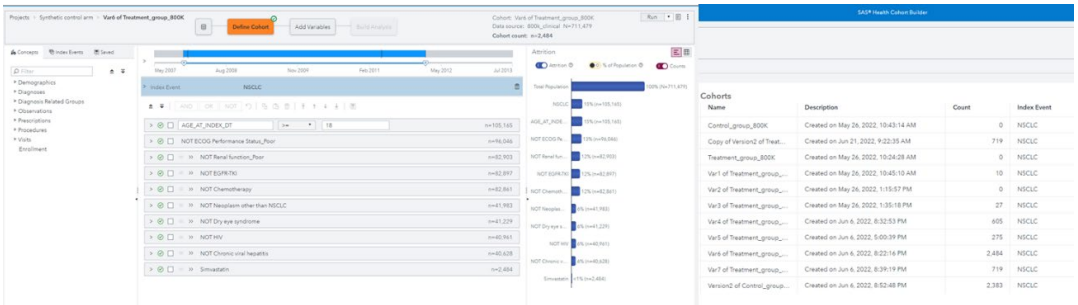

Figure 5. SAS Health Cohort Builder - build and manage cohorts

It is also important not to treat a RWE study project as a one-time effort. If the research team can manage the analytical assets accumulated through various study projects, they can **develop standardized analytics templates that can be reused commonly**, such as patient phenotype, validated definition of covariates, statistical analysis templates for safety or comparative studies, etc. It is a best practice to protect the organization's analytic IP and extend its existing library. With this practice, it also gives the research team the confidence that they are using programming code already validated by domain experts, and the statistical and research methods being used are appropriate to analyze RWD. Organizations can further establish best practices for studies that need to be conducted in routine, or simple studies can be covered by just using analysis templates, or complex studies needs customized programming, which are similar approaches taken by DARWIN EU.

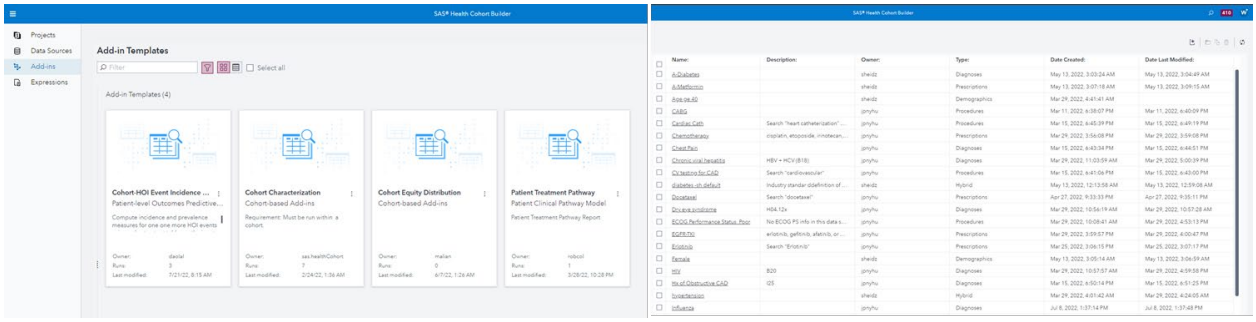

Figure 6. SAS Health Cohort Builder - manage standardized analytics assets

PROCESS – RELIABILITY & RELEVANCE

Besides DATA and METHOD, another important concept is the transparency and reproducibility of the whole RWE generation process. It is the key to gaining stakeholders' confidence. The analytics platform should be able to **provide a clear governance** on where the data came from, how the data were curated; as well as to **provide a full lineage from data to analysis**, including results, execution log and audit trial in order to increase transparency, traceability and reproducibility of the analyses. In addition, using a central environment to manage data, programs, metadata, and analytics assets is preferable from the perspective of traceability.

FEATURES OF A MODERN RWE ANALYTICS PLATFORM – OTHER CONSIDERATIONS

In this section, some other considerations that can help streamline the RWE generation process were discussed.

CLOUD NATIVE

A cloud-native platform which can deal with big data is recommended. The cloud-native platform provides access to the necessary computing power and storage while having the scalability and flexibility to resolve computational complexity. In addition, with the technology of in-database or in-memory parallel processing, the research team can deal with large volumes of data quickly in a high-performance environment. A cloud-native platform on a secure, GxP-qualified hosting environment can reduce the burden of IT, and the research team from every geographic location can access the analytics platform under strict access control rules.

COLLABORATION

A platform that enables full collaboration across different users with various skills from different divisions in the organization can maximize the value of RWD assets. The business analyst can use a data visualization tool to explore the data interactively, and a data scientist can work on a programming interface to analyze the data. With shared access and a low-code/no-code user interface, everyone can make data-driven decisions; and organizations can breakdown silos of data and analytics to enhance productivity.

CLASSICAL STATISTICS AND ADVANCED ANALYTICS

It is preferable that a platform can provide not only the classic statistical methods used in clinical trial analyses or health economic outcome research for submission purposes, but also provide the advanced analytics capability, such as AI/machine learning, computer vision, and natural language processing, for exploratory purposes. When it comes to advanced analytics for submission purposes in the future, the analytics platform should be able to provide the necessary governance and transparency of model training, scoring, and monitoring processes.

OPENNESS

It is preferable that the platform is open to multiple programming languages, such as SAS, R, Python, Julia, Lua, etc. With the flexible options, it allows people to work in the programming languages of their preferences, and it can benefit the organization in talent recruiting and talent retaining strategies. Another perspective of openness is being able to integrate with 3rd party applications. Also, having Open APIs to support access from outside systems could streamline business processes, from data management to building mobile applications for reporting.

SINGLE, END-TO-END

It is preferable to have a single, end-to-end analytics platform that can cover the whole analytics lifecycle. It means the users no need to waste time to manage and switch across multiple software for the tasks. All the tasks, from data preparation, data exploration, to data analyses and reporting, and modeling and model deployment, can be completed in one analytics platform, resulting in efficient governance and collaboration.

CONCLUSION

In conclusion, although the evidence generated from traditional RCTs is still considered the gold standard to prove the effectiveness and safety of a medical product, RWD/RWE has the potential to support the evidence generation process and to provide a more comprehensive aspect and more efficient approach to fulfill the unmet medical needs in the drug development life cycle.

Reliability, relevance and transparency of utilizing RWD, study design and analytical process are the key considerations for regulatory submission purpose. Having a trusted analytics platform in a GxP-compliant environment can help an organization leverage the value of RWD efficiently and confidently. For more usage of RWD and RWE for regulatory submissions, it is ideal to have an analytics platform that can provide sufficient computing power and efficient data management when dealing with big data from multiple sources. Also, it is preferable that the platform can provide a clear data provenance and systematic approach to control data quality, and a governed and cross-functional collaborative environment to increase speed to insights, which ultimately can improve health outcomes and patients' well-being.

REFERENCES

- US FDA. "Framework for FDA's RWE Program". 2018.
- HMA/EMA Big Data Steering Group. "Big Data Task Force final report". 2019.
- Nishioka, Kinue, et al. "Evolving acceptance and use of RWE for regulatory decision making on the benefit/risk assessment of a drug in Japan." Clinical Pharmacology & Therapeutics 111.1 (2022): 35-43.
- Feinberg, Bruce A., et al. "Use of real-world evidence to support FDA approval of oncology drugs." Value in Health 23.10 (2020): 1358-1365.
- Arondekar, Bhakti, et al. "Real-World Evidence in Support of Oncology Product Registration: A Systematic Review of New Drug Application and Biologics License Application Approvals from 2015–2020." Clinical Cancer Research 28.1 (2022): 27-35.
- Mary Kathryn Queen. "Be Prepared: An Introduction to SAS® Data Preparation". 2019
- Nancy Rausch. "What's New in SAS Data Management". 2020
- Mark L. Jordan. "Working with Big Data in SAS®". 2019
- SAS Help Center. Data Connectors
- dbInsight LLC. "SAS® Viya® and the cloud". 2020
- Stijn Rogiers et al. "PP19: Creative Data Science Using SAS Open Source Integration". 2021
- John Amrhein. Modeling Causal Effects Using SAS/STAT. 2021
- Douglas E. Faries et al. "Analysis of Observational Health Care Data Using SAS". 2014
- Douglas E. Faries et al. "Real World Health Care Data Analysis Causal Methods and Implementation Using SAS". 2020

ACKNOWLEDGMENTS

We would like to acknowledge the assistance of Dr. Mark Lambrecht (Director, EMEA & APAC Health and Life Sciences Practice) for valuable advice, input, and review of this paper.

CONTACT INFORMATION

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

Author Name: **William YuanHao Kuan**
Company: SAS Institute
Address: Roppongi Hills Mori Tower 11th floor, 6-10-1 Roppongi, Minato-ku, Tokyo 106-6111, Japan
Email: William.Kuan@sas.com
Web: https://www.linkedin.com/in/williamyhkuan/

Author Name: **Olivier Bouchard**
Company: SAS Institute
Address: Domaine de Grégy, Grégysur-Yerres Cedex, 77257 Brie Comte Robert Cedex, France
Email: Olivier.Bouchard@sas.com
Web: https://www.linkedin.com/in/olivierbouchard/