PharmaSUG 2025 - Paper RW-184

Addressing Challenges in Real-World Evidence Generation: The AI-SAS for Real-World Evidence Approach

Takuji Komeda, Shionogi & Co., Ltd.;
Yuki Yoshida, Shionogi & Co., Ltd.;
Yohei Komatsu, TIS Inc.;
Yoshitake Kitanishi, Shionogi & Co., Ltd.

ABSTRACT

Shionogi has developed a product, AI-SAS, that semi-automates programming tasks for clinical trials, leading to a 33% reduction in working hours. Our achievement, including AI-SAS, earned first place in the Innovative Problem Solver category at the 2024 SAS Customer Recognition Awards. Additionally, Shionogi is offering AI-SAS externally as part of our social contribution initiatives. In July 2024, the FDA issued guideline enabling the use of Real-world Evidence (RWE) in drug approval applications. Shionogi is expanding AI-SAS to generate RWE with plans to integrate generative AI technology. We develop a system using SAS Viya, which facilitates the implementation of machine learning and deep learning. To improve transparency, we identified key requirements: predefining analysis content in the protocol and statistical analysis plan (SAP) before conducting the analysis, creating analysis result reports, and ensuring consistency between document creation and analysis timings. To improve document creation efficiency, we use generative AI technology, which significantly assists researchers in drafting protocols and SAPs from their research questions. For executing analysis tasks, we use knowhow of AI-SAS, which semi-autogenerates programs from past specifications and mock-ups. These processes are recorded using GitHub in the system. This approach addresses both efficiency and transparency. The semi-automation process try to cover protocol, SAP, and specification creation, considering compliance with FDA guideline, and can achieve reduction in work time. The target audiences are persons who are interested in standardizing the process of analyzing RWDs and programmers who develop SAS macros efficiently.

INTRODUCTION

Shionogi & Co., Ltd. has been working on the development and utilization of AI-SAS (Official Name: AI-SAS Programmer System) to improve the efficiency of clinical trial analysis tasks. The concept of AI-SAS is to build predictive models of programs using past clinical trial data (text, numerical, image data) as training data. Subsequently, it reads the mock-up of the target clinical trial and semi-automatically generates the SAS program for analysis. This was presented at the SAS Global Forum in 2016 and 2017, achieving a reduction of 100 hours from the standard working time of 350 hours per trial in actual practice. Our achievements including AI-SAS won first place in the Innovative Problem Solver category of the 2024 SAS Customer Recognition Awards. Furthermore, Shionogi is also providing AI-SAS to those outside the company as part of its social contribution activities.

Data analysis in pharmaceutical companies covers a wide range of areas, and until recently, data analysis from clinical trials was the focus. However, there has been an increase in the utilization of RWD (Real-World Data). RWD refers to data relating to patient health status and healthcare provision that is collected routinely from various sources. Examples of RWD include data obtained from electronic medical records, medical claims data, product or disease registry data, data collected from other sources (such as digital health technologies) capable of providing health status information. The use of such data has clearly increased, with the number of papers utilizing RWD in Japan increasing from 4 in 2010 to 142 in 2020 according to PubMed (Figure 1 エラー! 参照元が見つかりません。) [1]. In 2024, a guidance titled "Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products"[2] was issued, enabling it to be used as evidence for drug review by the FDA, like clinical trials.

However, studies using RWD are sometimes criticized for their lack of transparency for various reasons, including:

- Acquiring large amounts of RWD and analyzing it through commercial license agreements without going through an ethical review board
- Not having to submit protocols or analysis plans in advance except when using them for applications

- Being able to complete all processes such as analysis, paper writing, and publication alone if access to data is available
- Most journals do not require the submission of the data used for analysis or its code. In other words, there is almost no third-party supervision, and it is impossible to deny the possibility that the research method was adjusted to obtain the results desired by the researchers.

The FDA guidelines also mention this. For example, it states, "For all studies using EHRs or medical claims data that will be submitted to FDA to support a regulatory decision, sponsors should submit protocols and statistical analysis plans before conducting the study." This statement is thought to be a comment made of concern for transparency.

Considering the current situation, we believe it is necessary to increase the efficiency of RWD analysis in a transparent flow to accelerate the creation of RWE (Real-World Evidence) and improve its quality. Therefore, we have initiated efforts to develop AI-SAS for RWE, which is an application and extension of AI-SAS for clinical trials.

In this paper, the overall of our new system, AI-SAS for RWE, is shown in detail. The target audiences are persons who are interested in the analysis of RWD, the process of the analysis of RWD and automation of the analysis process. Especially in this process, SAS Viya, Snowflake and GitHub are used so checking such kind of connections of systems is helpful for the persons.

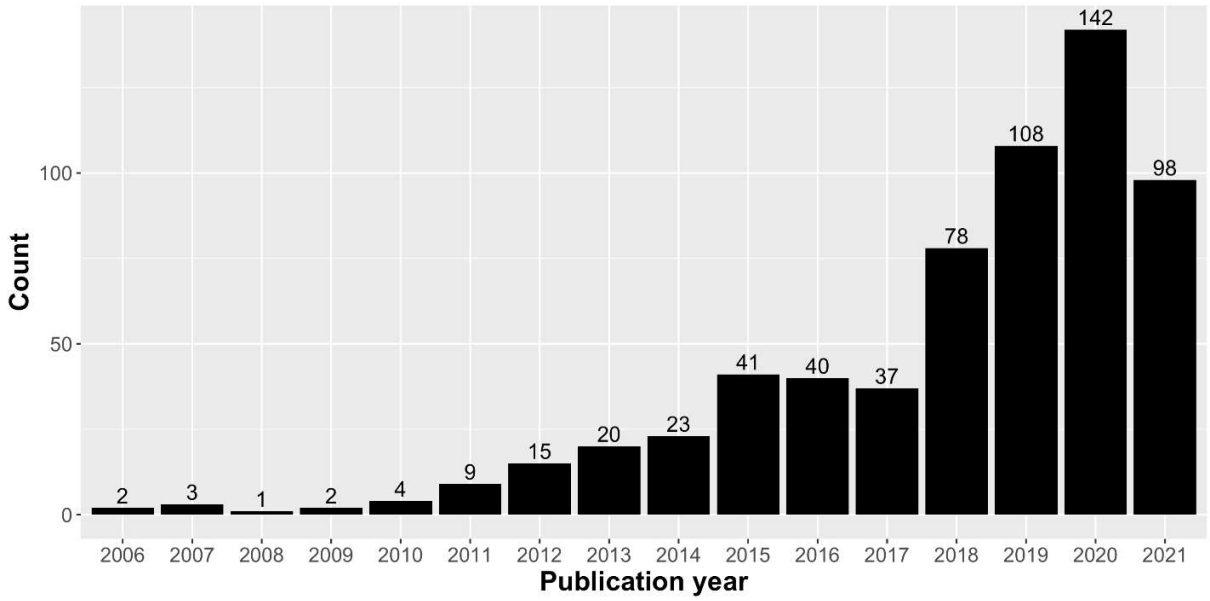

Figure 1 Number of publications related to RWD based on PubMed (January 2006 to July 2021) [1]

CONCEPT OF AI-SAS FOR RWE

As mentioned, the purposes of our system development are “transparency” and “efficiency.”

Typically, the process of RWD analysis is regulated as an SOP (Standard Operating Procedure) by each pharmaceutical company. Through this, companies ensure the reliability of the research results they conduct internally. However, since these regulations are internal, the research implementation process is sometimes criticized for lacking transparency.

The reason for this is that there is minimal third-party oversight, making it impossible to completely rule out the possibility that researchers may have adjusted their research methods to obtain favorable results [3]. (Figure 2)

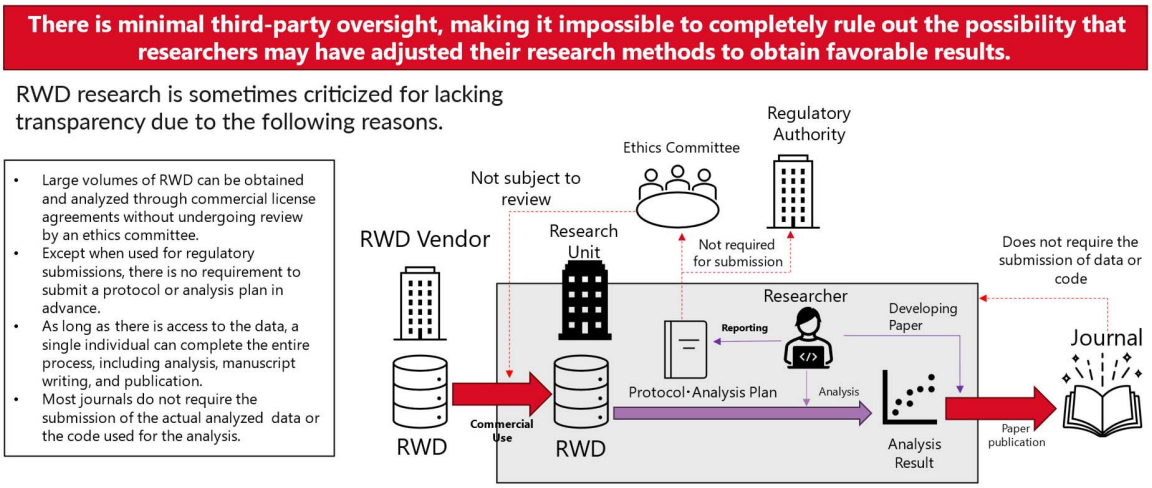

Figure 2 Transparency Issues in Research Using RWD

To examine efficiency improvements, we analyzed the causes of inefficiencies in past research conducted within the company. We then compared these results with the conditions of clinical trials that had already achieved efficiency improvements using AI-SAS. The identified causes were classified into the following categories: raw data format, data scale, types and formats of analysis materials, and analysis methods. The details are as follows:

- Raw data format: The format differs by vendor, so the same process cannot be applied to the entire RWD.
- Data scale: A significant number of resources are allocated to creating analytical variables from massive records for the creation of the Analytical Data Set (ADS).
- Types and formats of analysis materials: There are no industry standards, so data for machine learning to learn is not organized.
- Analysis method: The invention and implementation of analysis methods that deal with confounding factors is necessary.

Therefore, we have proceeded with the approach of "RWD standardization (Common Data Model)" to organize the raw data formats from each vendor and "semi-automatic RWD analysis" to improve efficiency and transparency of the entire analysis process, including the creation of ADS. In other words, by standardizing RWD and semi-automating the analysis process and automatic recording activities including the documentations, we aimed to improve the efficiency of RWD analysis in a transparent flow and accelerate the creation and quality improvement of RWE (Figure 3).

We are also establishing an environment where access rights and analysis records can be managed through external systems such as **Snowflake** and **GitHub**.

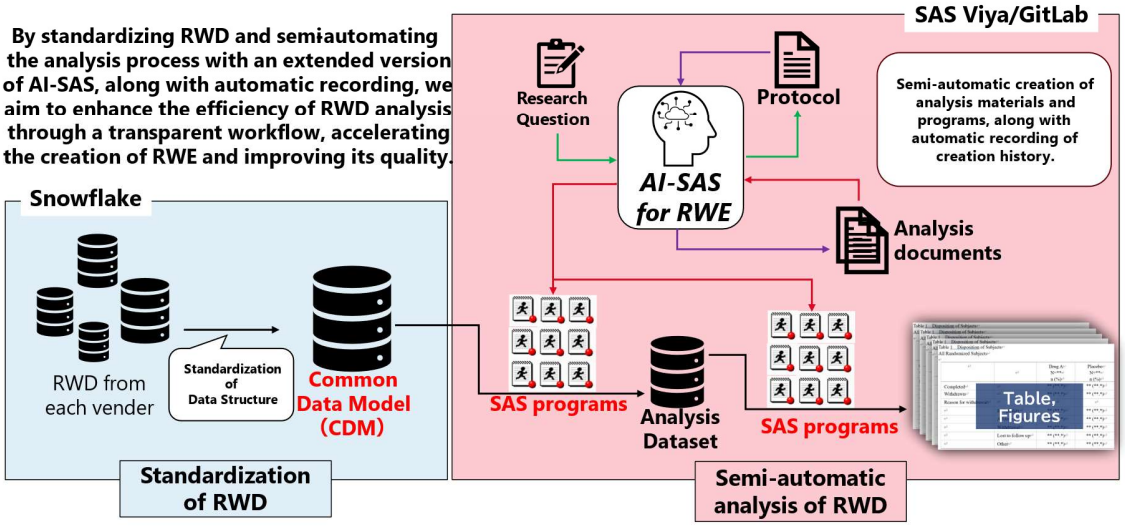

Figure 3 Concept of AI-SAS for RWE

SEMI-AUTOMATED APPROACH FOR RWD ANALYSIS

In the previous section, we explained the concept and overall system of AI-SAS for RWE. Here, we will describe the details of the automated analysis process within this system.

This project is anticipated to be developed in three steps:
Step 1: Creation of basic analysis materials and semi-automation of analysis
Step 2: Further efficiency improvement and expansion of the automatic generation range by utilizing AI
Step 3: Improvement of usability through application (Figure 4).

The details of each step are described below.

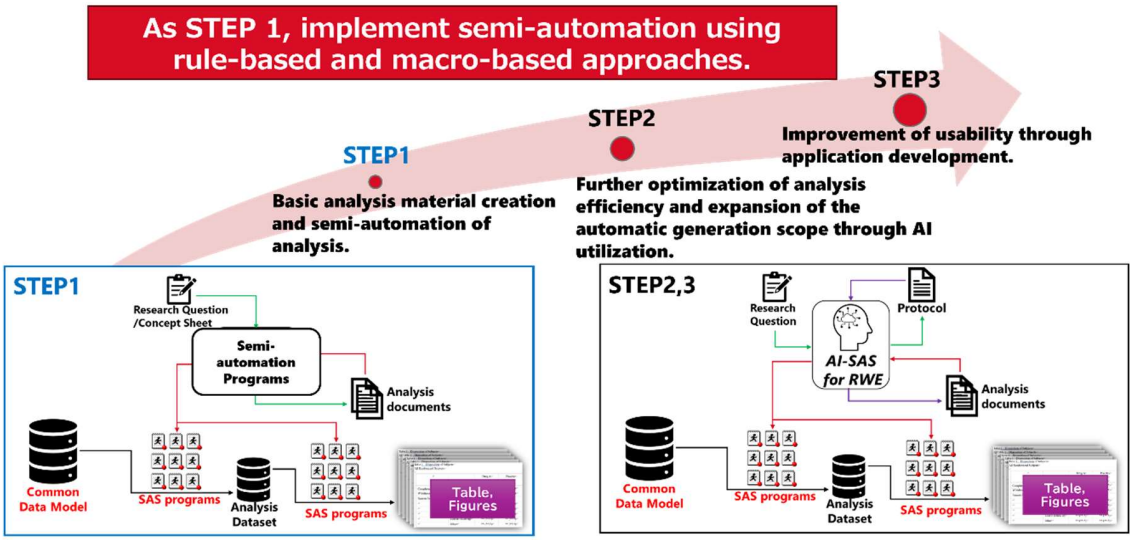

Figure 4 Semi-automated Approach for RWD Analysis

Regarding Step 1, firstly, a concept sheet is used to input PECO (Patient, Exposure, Comparison, Outcome) along with background factors, analysis methods, and other relevant information. Based on this input, ADS Spec and TLF Spec are generated semi-automatically. During the creation of ADS Spec, a knowledge DB that stores variable definitions is referenced. In RWD analysis, defining variables is a crucial and time-consuming task that also requires medical expertise. Therefore, at Shionogi, past variable definitions from the company's own research, as well as definitions from previously published papers in similar disease areas, are stored in the database and referenced as needed. Using these resources, SAS programs are also generated semi-automatically. However, since this process is not fully automated—allowing human intervention for modifications and adjustments—it is considered "semi-automatic."

From the Common Data Model, ADS and TLF are semi-automatically created. Additionally, to ensure transparency, both the analysis conditions and results recorded in the concept sheet are automatically compiled into a single report file. All these processes are executed on SAS Viya, which facilitates the implementation of machine learning, deep learning, and large language models (LLMs). Furthermore, all processes are configured to be automatically reflected in GitHub.

Through this approach, both efficiency and transparency in RWD analysis are achieved. (Figure 5)

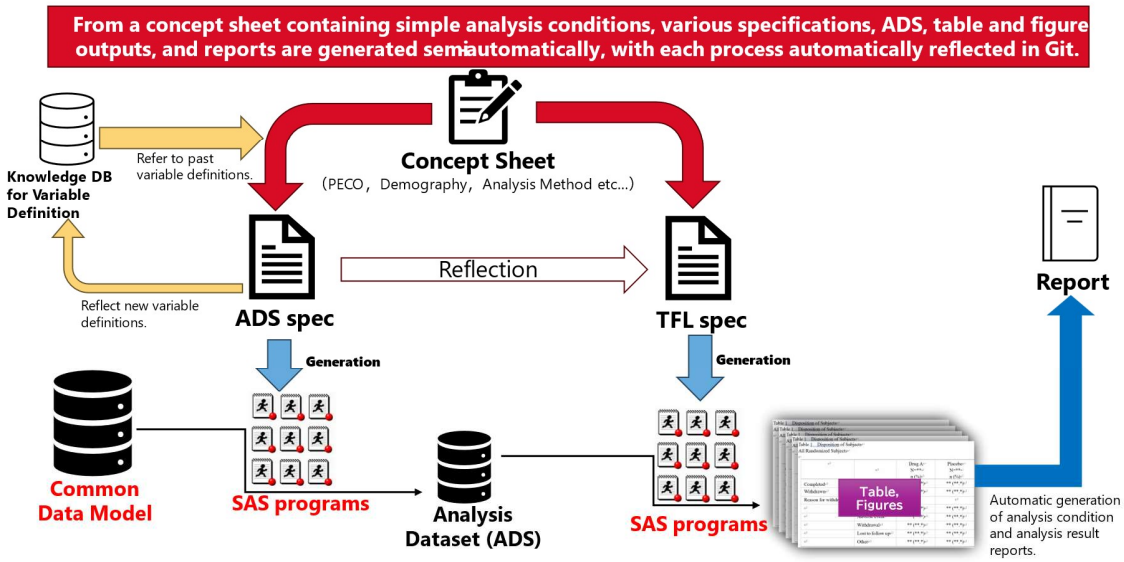

Figure 5 Image of Implementation of STEP 1

Step 2 is currently under development. We plan to include generative AI technology, which efficiently creates Protocols, SAPs, etc. For example, the FDA's guidance [2] includes the following description regarding items to be included in the protocol, SAP: “The protocol and the statistical analysis plan should be developed and based on an understanding of reasons for the presence and absence of information. Descriptive analyses should be included to characterize the missing data. Assumptions regarding the missing data (e.g., missing at random, missing not at random) underlying the statistical analysis for study endpoints and important covariates should be supported and the implications of missing data considered in the design and analysis of the study. Sensitivity analysis should be conducted to evaluate the robustness of findings.” Regarding these, considerations are necessary when wording the selection of the method and the details of the selected method. By using generative AI technology, this step can be simplified significantly. Moreover, by using a system that has implemented it, it is possible to make the process transparent, meaning that it is possible to leave a record.

CONCLUSION

In this paper, the increasing role of RWD in the pharmaceutical industry was described, driven by the growing availability and improving quality of RWD. Despite this expansion, challenges related to resource allocation and the quality of generated outputs were identified—issues that are likely to be shared not only by Shionogi but also by other pharmaceutical companies.

To address these challenges, AI-SAS for RWE, a system developed by Shionogi, was introduced. The key aspects of this system are automation and transparency. Through automation, the time required for programming, which previously took several months, has been reduced to just a few minutes, although within a limited input and output framework. Regarding transparency, a transition was demonstrated from manual SOP-based process management to an automated process tracking system.

It should be emphasized that these achievements remain at the STEP 1 stage. In STEP 2, efforts will be made to move beyond efficiency improvements and develop mechanisms that add value to the research

process, such as AI-driven research question formulation. By continuing to evolve alongside advancements in AI, AI-SAS for RWE is expected to contribute to enhancing the value of RWD utilization.

REFERENCES

[1] Zhao Y, Tsubota T. 2023 Feb 14 “The Current Status of Secondary Use of Claims, Electronic Medical Records, and Electronic Health Records in Epidemiology in Japan: Narrative Literature Review.” *JMIR Med Inform.* 11:e39876. doi: 10.2196/39876. PMID: 36787161; PMCID: PMC9975931.

[2] U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Oncology Center of Excellence (OCE). Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products. July 2024. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory

[3] Dagenais S, Russo L, Madsen A, Webster J, Becnel L. Use of Real-World Evidence to Drive Drug Development Strategy and Inform Clinical Trial Design. Clin Pharmacol Ther. 2022 Jan;111(1):77-89. doi: 10.1002/cpt.2480. Epub 2021 Nov 28. PMID: 34839524; PMCID: PMC9299990.

CONTACT INFORMATION

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

Takuji Komeda

Shionogi & Co., Ltd.

takuji.komeda@shionogi.co.jp