Analyzing Clinical Trial Data from Multiple Studies Using SAS® Real World Evidence

Jay Paulson, Data Scientist, Health and Life Sciences R&D, SAS Institute

David Olaleye, Senior Manager, Health and Life Sciences R&D, SAS Institute

ABSTRACT

This paper explains how SAS® Real World Evidence solution makes use of data sets obtained from multiple clinical trial data sources to derive intelligence beyond the scope of a typical clinical trial. By using Prostate Cancer clinical trial data from Project Data Sphere as example, this paper explains the concept of Workspaces, Index Cohorts, Population Cohorts and Expressions. Then the paper explains how users can create Workspaces, Index Cohorts, Population Cohorts and Expressions within SAS® Real World Evidence framework to perform analysis. The paper also explains how Index and Population cohorts will be used to perform analyses, derive insights and answer critical research questions.

INTRODUCTION

In clinical research, data from completed clinical trials have been protected and used exclusively for their own business purposes by both academic researchers and pharmaceutical companies primarily for competitive reasons although this data provides a wealth of information and insights into several diseases for which no cure have been found. The idea of clinical trial data sharing is moving beyond past hurdles with the emergence of several sharing models like Project Data Sphere (PDS). The availability of sharing models combined with the power of SAS® Real World Evidence provide researchers unprecedented access to a wealth of information. After downloading data from multiple clinical trials the data needs to be mapped and integrated to make it ready for SAS® Real World Evidence. Once the data is available to SAS® Real World Evidence researchers can leverage the power of SAS technologies to analyze the data.

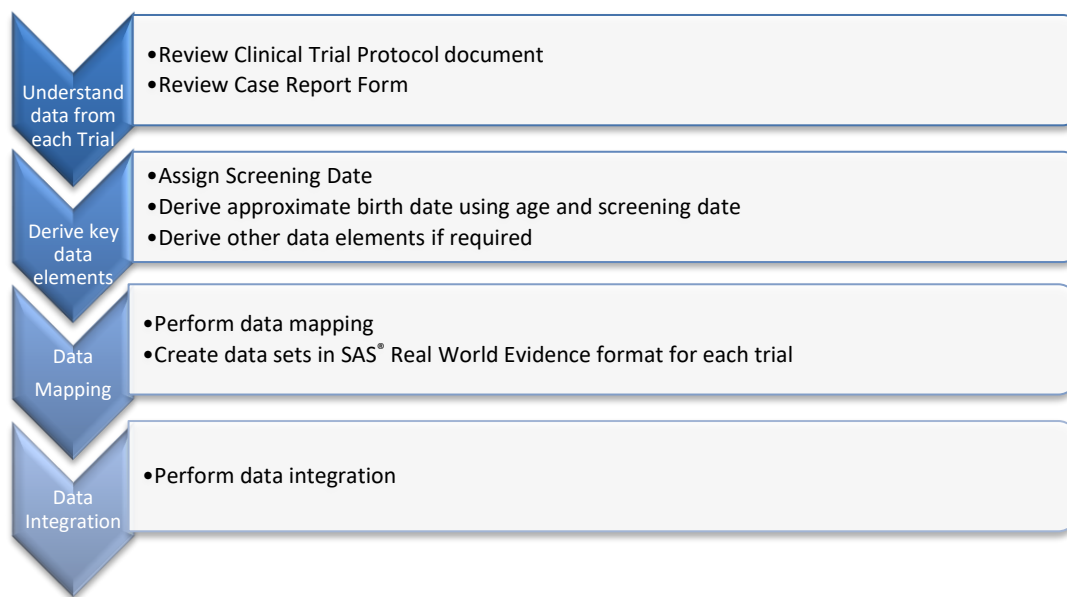## DATA FROM PROJECT DATA SPHERE

For this paper, we have used Prostate cancer trial data from Project Data Sphere (PDS). PDS is an example of an active open-source data-sharing model. PDS is a free digital library-data laboratory that was launched in 2014 as an independent, nonprofit initiative of the CEO Roundtable on Cancer, which was founded in 2001 to develop and implement initiatives that reduce the risk of cancer, enable early diagnosis, facilitate access to the best available treatments, and hasten the discovery of new and more effective anticancer therapies. A Web-based platform for accessing open-source data was developed for PDS by SAS Institute. Using this website, researchers can download, share, integrate, and analyze patient-level data. Data contributors are provided access to deidentification and data-standardization protocols, as well as to templates of legal agreements, including standardized data-sharing and online-services user agreements. Users of the site have access to analytic tools developed by SAS Institute. Anyone interested in cancer research can use the website, if they register and agree to a responsible-use attestation. PDS is funded by the engagement of a wide range of stakeholders that together ameliorate the burden of securing adequate funding from a single organization or institution (Bertagnolli et al., 2017).

PROSTATE CANCER TRIAL DATA FROM PROJECT DATA SPHERE

Prostate Cancer data from 13 clinical trials were downloaded from Project Data Sphere website as SAS data sets into 13 different folders. These data sets were checked for data quality before mapping and integration.

MAPPING PROJECT DATA SPHERE DATA TO SAS® REAL WORLD EVIDENCE

As the data sets from multiple clinical trials have variations in structure and format based on the study design across multiple trials, a thorough review of relevant documents including case report form and clinical trial protocol document for each trial was required. As it was virtually impossible to collect some key data elements required in SAS® Real World Evidence (e.g., date of birth of the subject) these elements had to be derived. A summary of different steps involved is given in Figure 1.

Understand data from each Trial	• Review Clinical Trial Protocol document • Review Case Report Form
Derive key data elements	• Assign Screening Date • Derive approximate birth date using age and screening date • Derive other data elements if required
Data Mapping	• Perform data mapping • Create data sets in SAS® Real World Evidence format for each trial
Data Integration	• Perform data integration

Figure 1. Steps in mapping Project Data Sphere data to SAS® Real World Evidence

INTRODUCTION TO SAS® REAL WORLD EVIDENCE

SAS® Real World Evidence lets researchers and analysts harness the power of real-world and health care data to gain insights. Data can come from multiple data sources and vendors. Workflows have been created to walk users through the process of defining and building a cohort, which is a list of patients that meet specific criteria. SAS® Real World Evidence main page is shown in Figure 2.
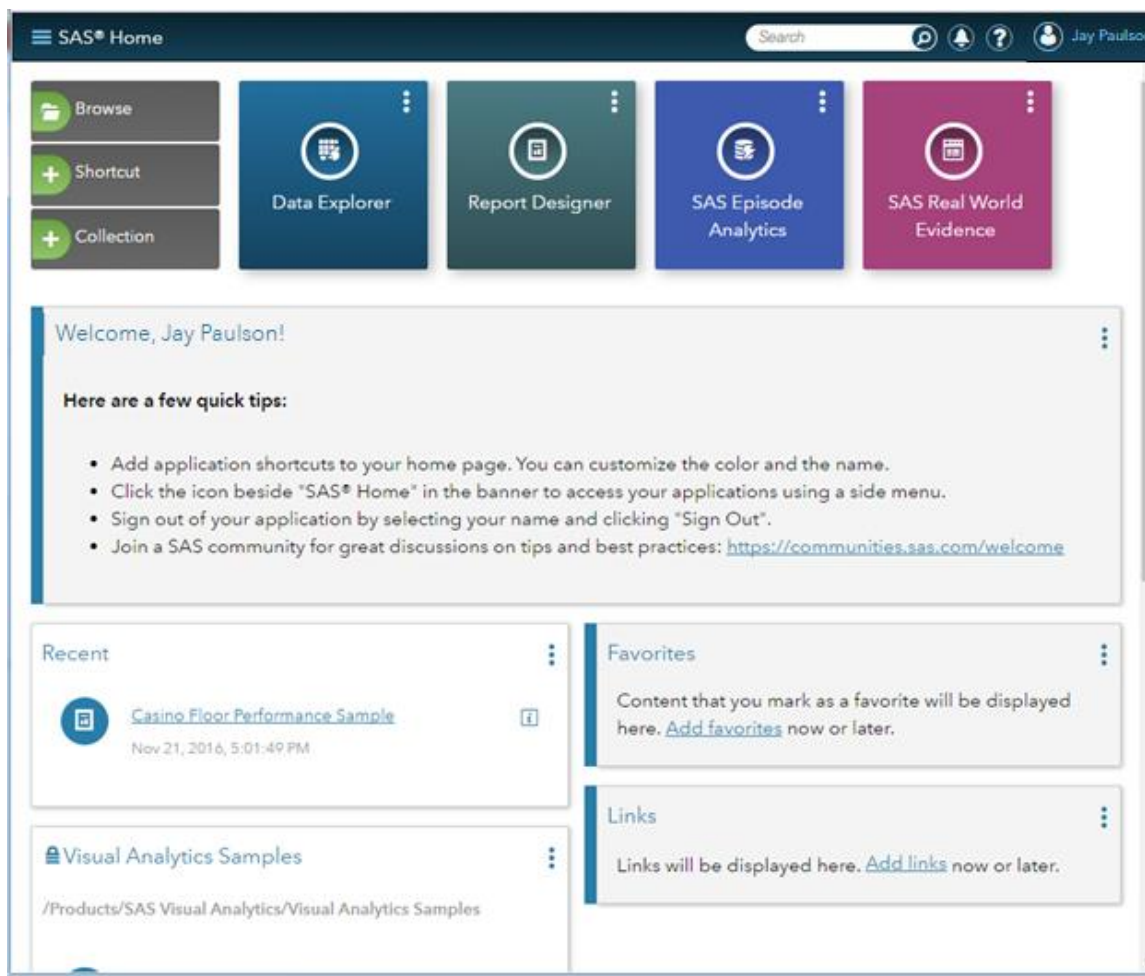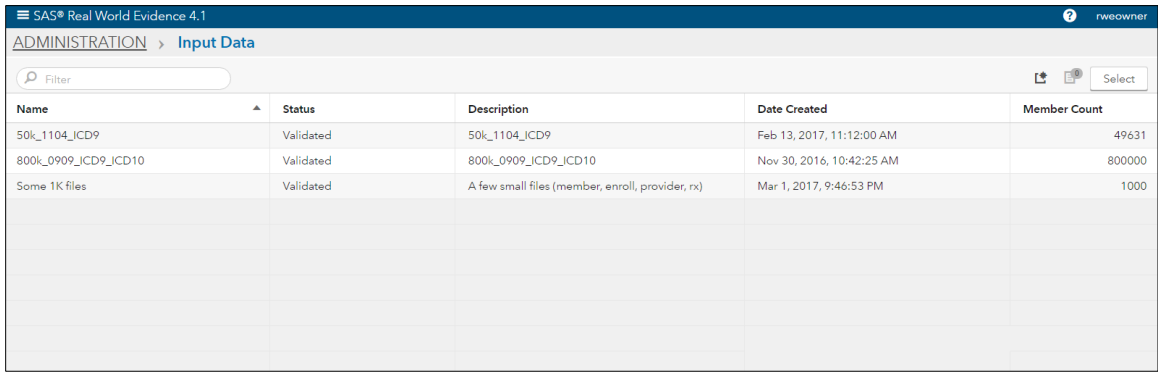
Figure 2. SAS® Real World Evidence main page

UPLOADING DATA INTO SAS® REAL WORLD EVIDENCE

SAS® Real World Evidence provides a tool for uploading data. This tool can be accessed from SAS® Real World Evidence administration module. Figure 3 shows the input data administration screen. From the administration screen a user can bring up the screen to define a new input data source. This screen allows the user to pick raw data files from a hard drive for uploading to SAS® Real World Evidence as shown in Figure 4.
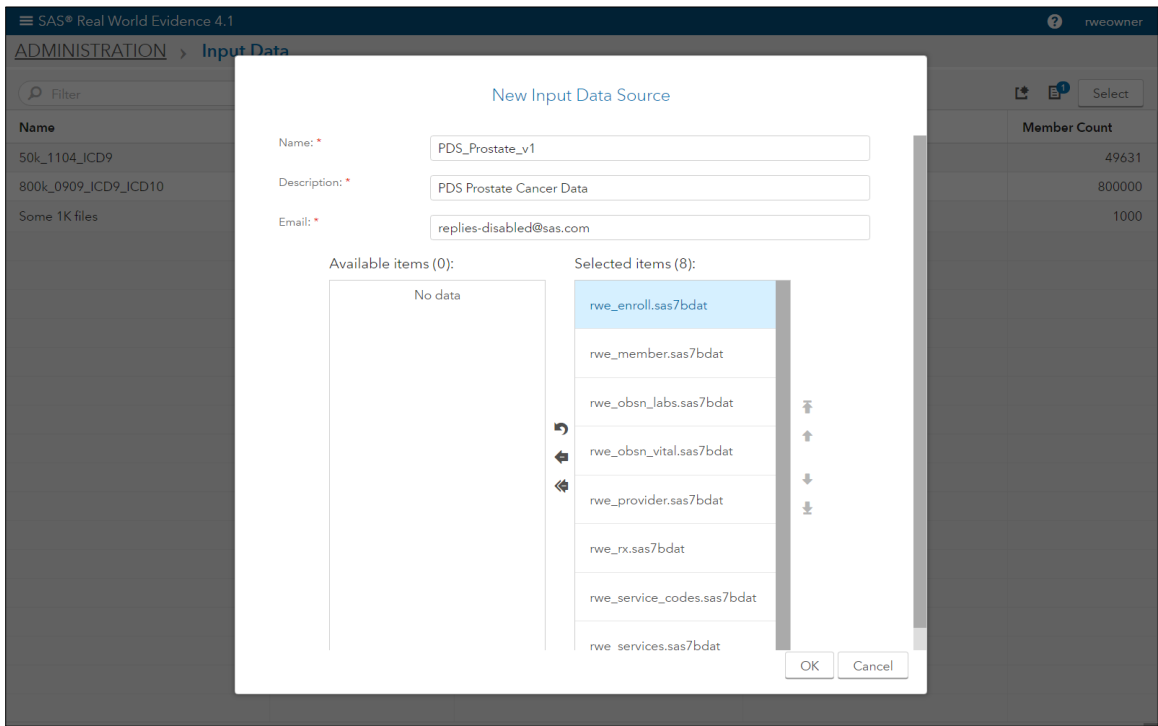
After the data is validated and uploaded successfully, it will get listed on the input data administration screen. For example, we created the data source named "PDS_Prostate_v1" in SAS® Real World Evidence with Prostate cancer data sets as shown in Figure 5. This data source will be used as example throughout this paper.

Figure 3. Administration screen for loading data to SAS® Real World Evidence
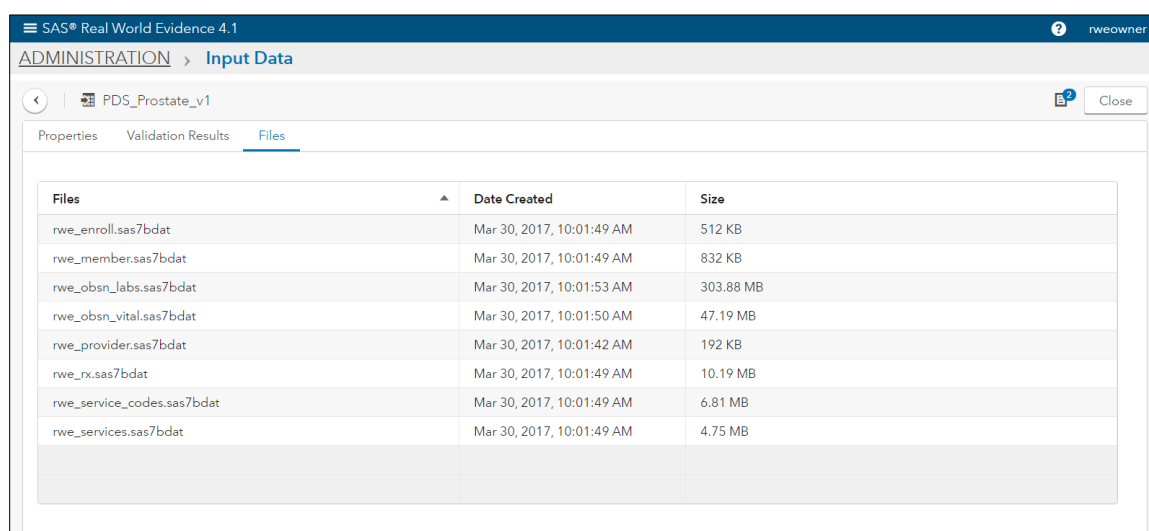
Figure 4. Defining a new input data source in SAS® Real World Evidence

The input data administration screen (Figure 5) shows the new data with the status "Validated". The details of uploaded data sets can be viewed as shown in Figure 6.

Figure 5. Validated data in SAS® Real World Evidence

Figure 6. Details of validated data

CONCEPT OF WORKSPACES

Workspaces contain the jobs that create cohorts, plus the cohorts themselves, along with any reports and data that are created for the cohort and modeling jobs. For example, Figure 7 shows five workspaces.
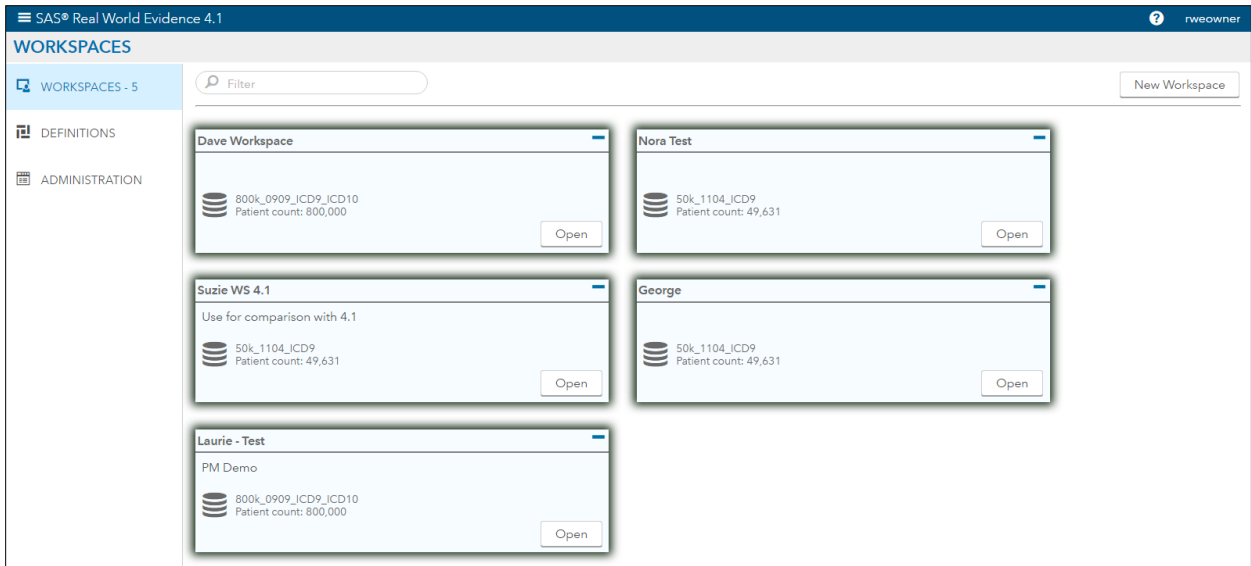
Figure 7. Workspaces in SAS® Real World Evidence

CREATING A WORKSPACE IN SAS® REAL WORLD EVIDENCE

A new workspace called "PDS_WS1" was created for our analysis using the data source "PDS_Prostate_v1" as shown in Figure 8.
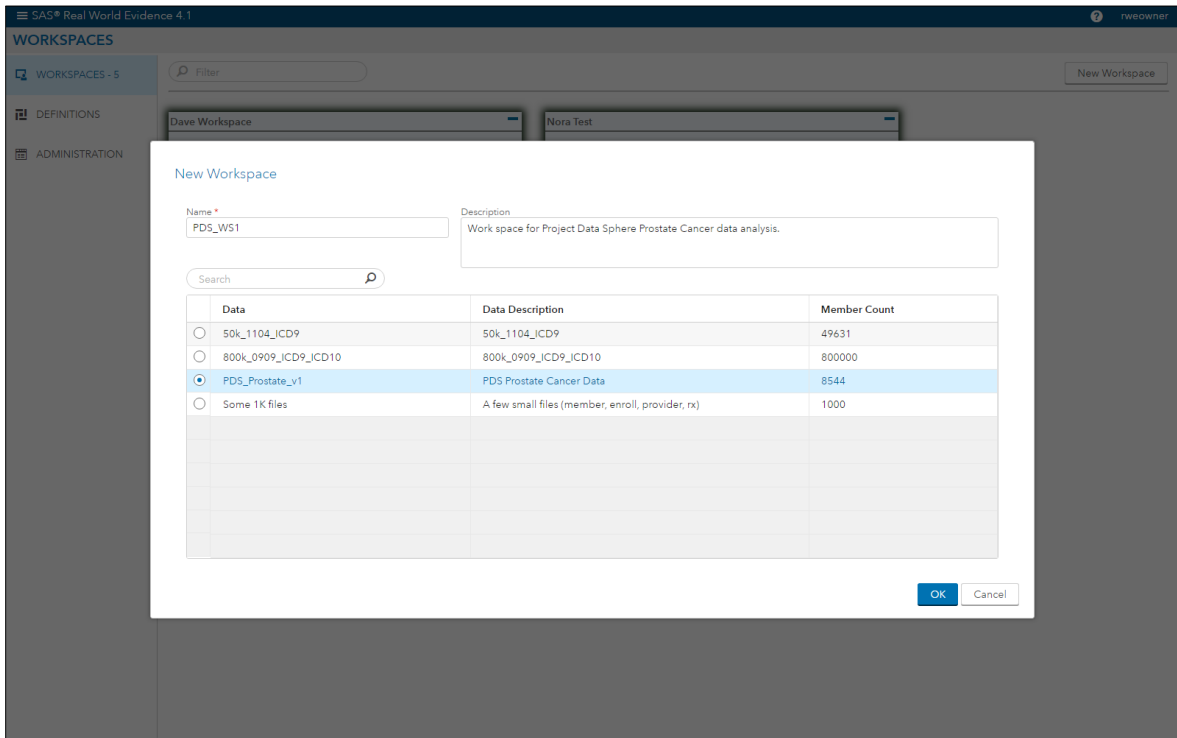

Figure 8. Creating a new workspace in SAS® Real World Evidence

ANALYZING WORKSPACE PROFILE

Right after the workspace is created, SAS® Real World Evidence provides a wealth of information about the workspace including workspace profiling report (Figure 9) and a snapshot of data sets (Figure 10). For example, the age distribution graph in Figure 9 indicates that the highest number of Prostate cancer patients are in the age group "65-<70" with a total of 2139 patients. As indicated in the pie chart (bottom, middle) about 20% of the subjects died during trials.

Figure 9. Analyze workspace profile

Snapshots of all data sets give users a feel for the data including key variables, data quality etc. A snapshot of labs data is given in Figure 10.

Figure 10. Details of data sets in the workspace

CONCEPT OF COHORTS

A cohort is a group of subjects being analyzed. There are two types of cohorts: Population Cohort and Index Event Cohort.

POPULATION COHORT

A Population Cohort enables users to identify patients who meet selected criteria within a specified study period.

INDEX EVENT COHORT

An Index Event Cohort enables users to identify patients that meet the following criteria:

- have an index event within the identification window. The identification window is the period during which the index event must occur for the patient to qualify for the cohort.

- meet indicated enrollment requirements, if applicable

- have data for the entire study period

- meet indicated criteria for additional parameters, if specified
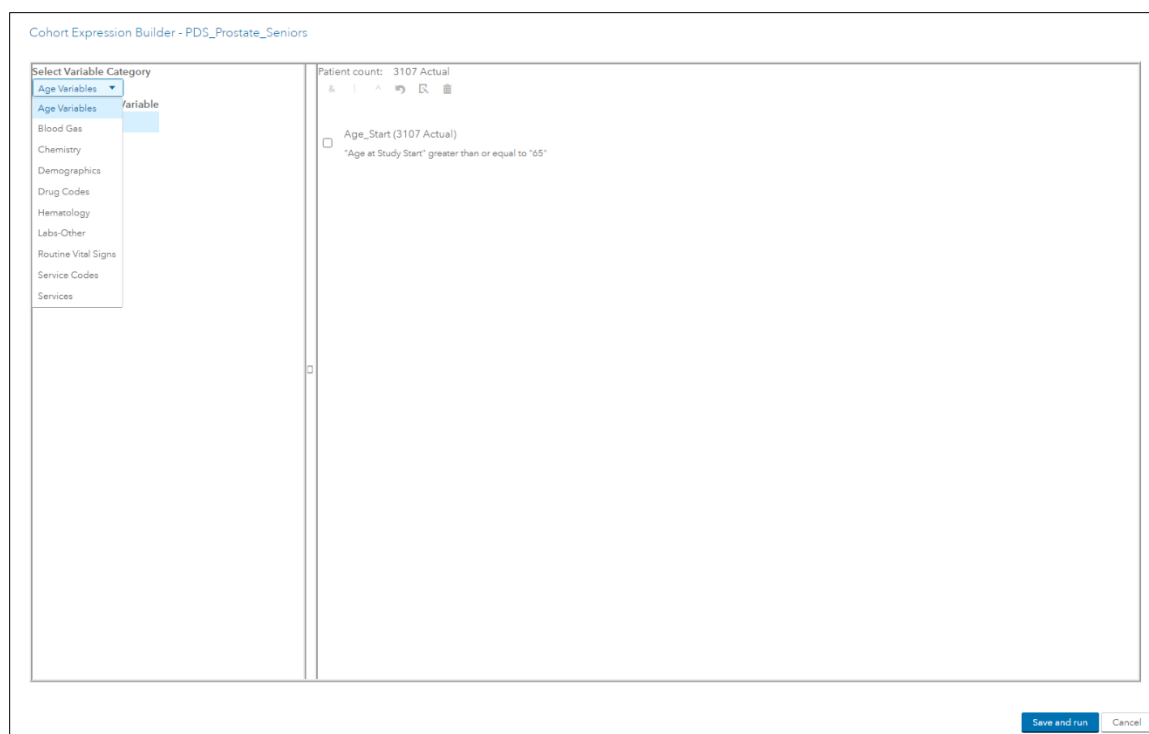
CREATING A POPULATION COHORT

A population cohort was created based on the PDS Prostate Cancer data for the subjects who are age 65 or above. The objective of this cohort is to analyze the characteristics of Prostate Cancer patients who are senior citizens. Figure 11 shows how the population cohort was created.

**Figure 11. Creating a Population Cohort**

After a cohort is defined, the user gets the opportunity to narrow down the results further by providing specific details about the cohort including demographic details, lab tests, vital signs, drug code etc. using Expression Builder functionality as shown in Figure 12.

Figure 12. Using expressions

After a population cohort is created, details about the cohort is displayed on the cohort main screen (Figure 13). For example, the Prostate Cancer cohort has 3107 patients and study period ranges from June 11, 1994 to January 18, 2013.
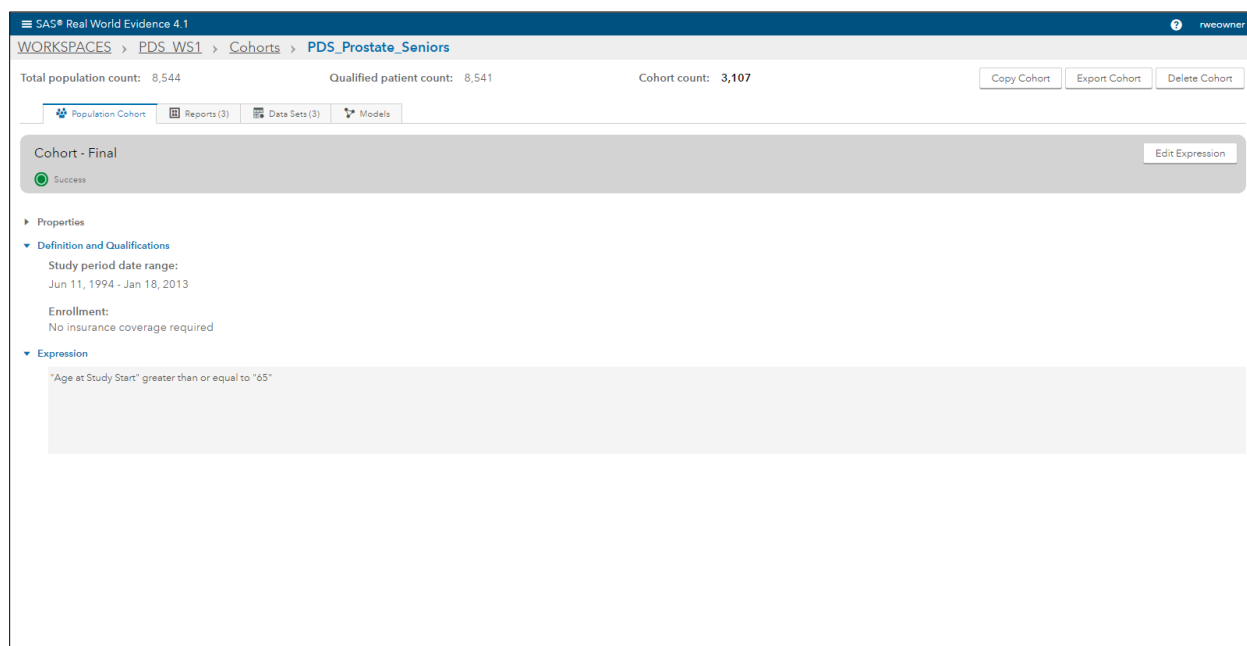
Figure 13. Population Cohort

A detailed analysis of cohort demographics can be performed using the cohort summary screen (Figure 14). Details about different data sets in the cohort can be obtained from the data sets tab (Figure 15).
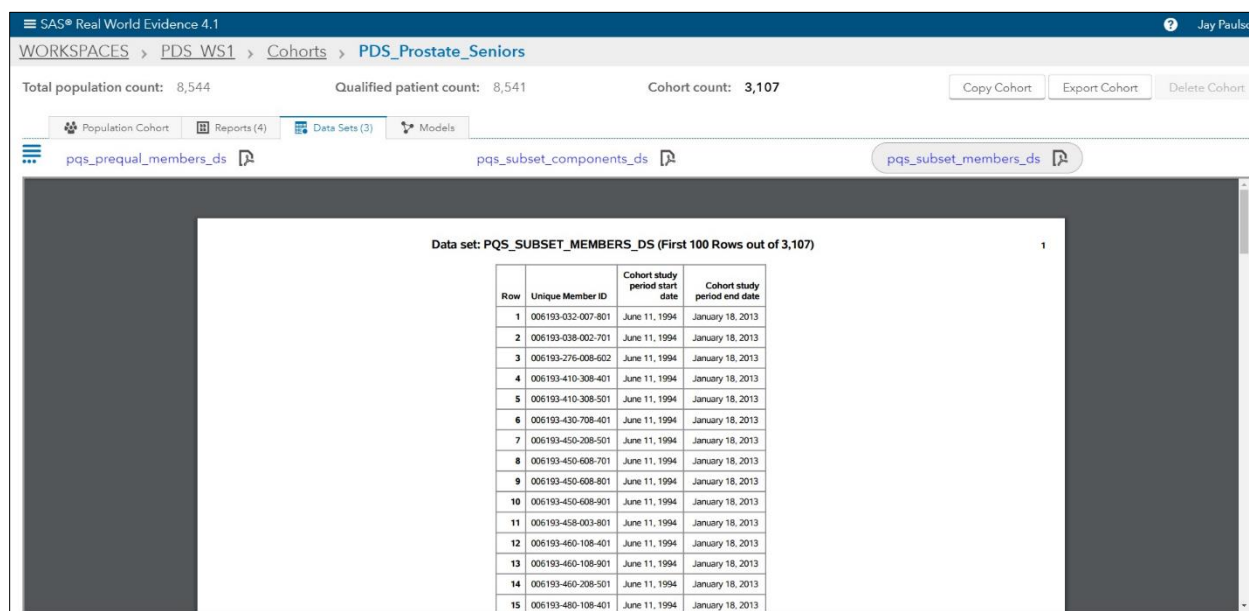

Figure 14. Cohort Summary

Figure 15. Cohort data sets

ANALYTICAL MODELS

Analytical models in SAS® Real World Evidence can be used to perform in-depth analysis of cohort data. One example is the Cohort Characterization Model that provides an analysis of key characteristics of a cohort including service activity, medications, and diagnoses.

COHORT CHARACTERIZATION

For the PDS Prostrate data we analyzed the demographics using the charts in Figure 16. Top drugs taken by these subjects based on frequency were analyzed using the table in Figure 17. The #1 drug was Acetylsalicylic acid, an analgesic. Top adverse event based on frequency for this cohort was found to be Fatigue followed by Constipation and Nausea (Figure 18).
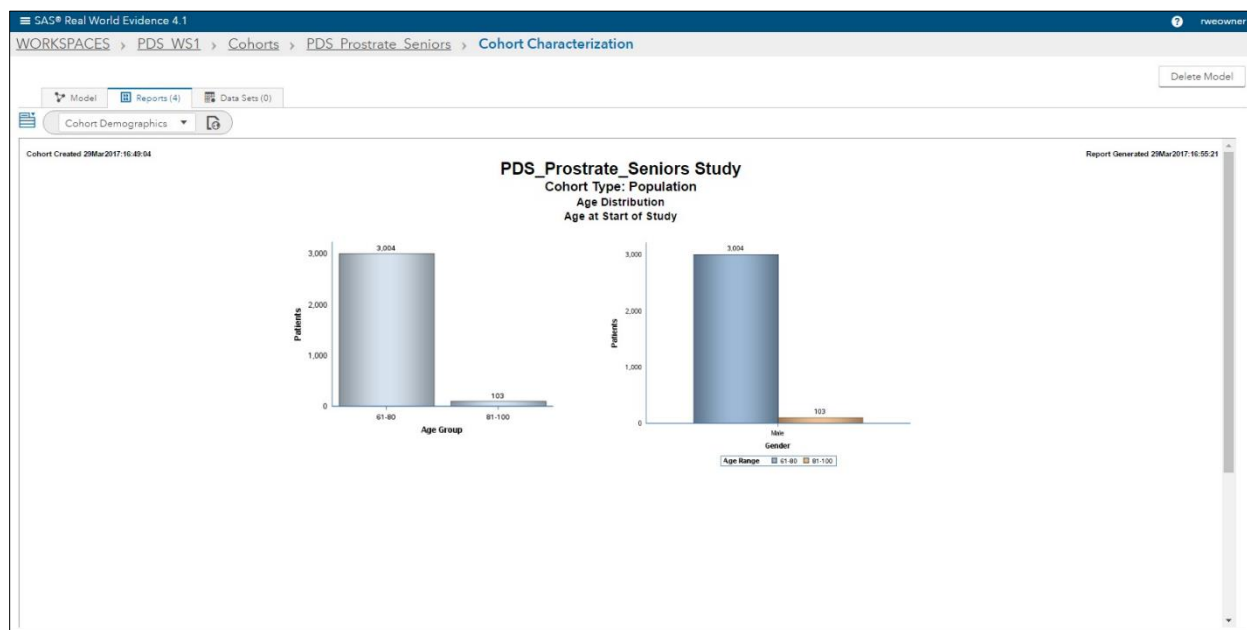

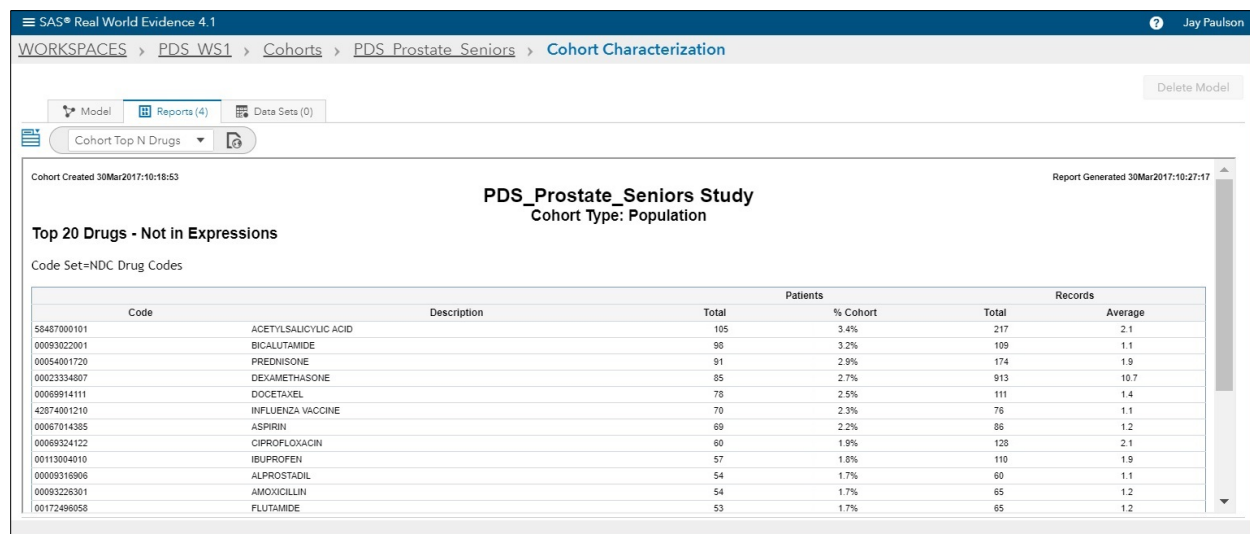**Figure 16. Cohort Characterization-Demographics**

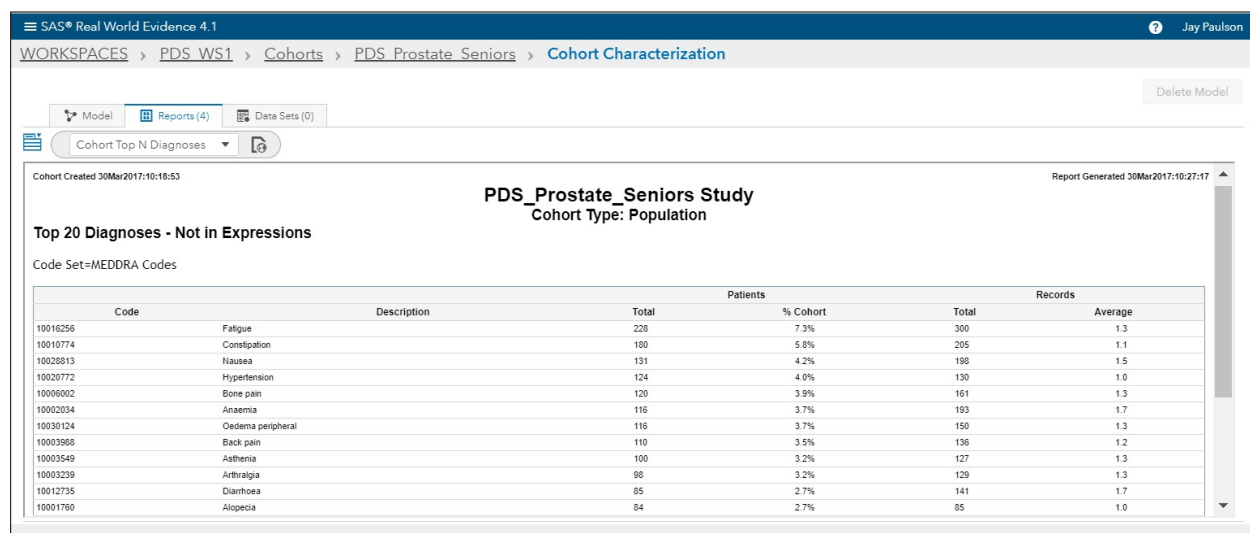**Figure 17. Cohort Characterization-Top Drugs based on frequency**

Figure 18. Cohort Characterization-Top Adverse Events based on frequency

CONCLUSION

This analysis demonstrates the power of SAS® Real World Evidence in analyzing and gaining insights beyond the scope of a single clinical trial. Although the definition of Real World Evidence is still evolving, most proponents associate Real World Evidence with data that is derived from medical practice among heterogeneous sets of patients in real-life practice settings, such as insurance claims data and clinical data from electronic health records. SAS® Real World Evidence has the power and versatility to make this possible.

REFERENCES

Monica M. Bertagnolli, M.D., Oliver Sartor, M.D., Bruce A. Chabner, M.D., Mace L. Rothenberg, M.D., Sean Khozin, M.D., M.P.H., Charles Hugh-Jones, M.D., David M. Reese, M.D., and Martin J. Murphy, D.Med.Sc., Ph.D. March 23, 2017. "Advantages of a Truly Open-Access Data-Sharing Model." The New England Journal of Medicine, 10.1056/NEJMsb1702054

Thomas E. Hubbard, Rebecca Paradis. September 2015. "Real World Evidence: A New Era for Health Care Innovation", The Network for Excellence in Health Innovation. http://www.nehi.net

SAS Institute Inc. 2016. SAS® Real World Evidence 4.1: Data Dictionary. Accessed March 30, 2017. https://support.sas.com/documentation/prod-h/rwe/4.1/rwedd.pdf

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

Jay Paulson
SAS Institute
jay.paulson@sas.com
http://www.sas.com