Measuring Study Potential Through the Use of Data Diagnostics

Clair Blacketer[1,2], Frank DeFalco[1]

[1]Janssen Research & Development, Raritan, NJ, [2]Eramus MC, Rotterdam, NL, [3]Department of Biomedical Informatics, Columbia University, New York, NY, USA

Background

The 2023 OHDSI Save Our Sisyphus (SOS) Challenge kicked off in March 2023 with the intent to demonstrate how to run network studies using OHDSI tools and methods across a federated network of databases(1). During the challenge the Data Diagnostics tool was introduced to assess database feasibility. Data Diagnostics uses pre-computed database-level summary statistics to determine if a database is deemed fit to answer any number of clinical questions. It takes as an input user-defined values for age, minimum person observation time, calendar dates, required domains of coverage (conditions, procedures, etc.), target concepts, comparator concepts, indication concepts, and outcome concepts among others. Using only the summary statistics it returns a report detailing which databases in a network meet each of the criteria the user defined(2). This analysis examined one of the four SOS challenge studies and how well data diagnostics performed. We compared the output of data diagnostics for the study "Risk of kidney failure associated with intravitreal anti-vascular endothelial growth factor (anti-VEGF)" with the output of study diagnostics to evaluate how they align and if data diagnostics was able to accurately identify databases that are and are not fit to answer the individual study questions.

Methods

The anti-VEGF study was initiated with the purpose to estimate the risk of kidney failure after intravitreal anti-VEGF, comparing the drugs ranibizumab, aflibercept, and bevacizumab. Prior to requesting OHDSI data partners run the study, data diagnostics was run to identify databases that would be the best fit. The data diagnostics inputs were defined as follows for the comparison of aflibercept and bevacizumab:

Analysis Name = "aflibercept vs. bevacizumab for blinding diseases with ESRD outcome"
Minimum Age = 18
Maximum Age = *No Restrictions*
Genders = Male, Female
Races = *No Restrictions*
Ethnicities = *No Restrictions*
Study Start Date = *No Restrictions*
Study End Date = *No Restrictions*
Required Observation Days per Person = 365
Required Data Domains = Conditions, Drugs
Target = aflibercept
Comparator = bevacizumab
Outcome = End Stage Renal Disease

Indication = Blinding Diseases

The target and comparator were then switched in and out between the three drugs identified for a total of three comparisons.

A subset of the potential database collaborators that were identified as being fit to participate ran the full study which included cohort generation, cohort diagnostics, characterization, prediction, and estimation. Detailed methodology on each piece can be found here(3). In the present analysis we followed the databases, characterizing which ran the three study questions, and which passed study diagnostics. The names of the data partners are masked to align with data governance requirements.

Results

A total of 31 databases were assessed for the potential to answer the study questions comparing the three anti-VEGF medications for the outcome end stage renal disease. For the comparison aflibercept to bevacizumab, 14 were determined to have all the necessary elements required for the study. One database did not pass data diagnostics but ran the study package and is also included in the analysis. Table 1 shows by data partner (DP) which diagnostics passed and which databases ultimately contributed results.

Table 1. Data and Study Diagnostic Results by Data Partner for the comparison of aflibercept and bevacizumab for the outcome end stage renal disease

	Passed Data Diagnostics	Ran Study Diagnostics	Passed Study Diagnostics
DP1			
DP2			
DP3			
DP4			
DP5			
DP6			
DP7			
DP8			
DP9			
DP10			
DP11			
DP12			
DP13			
DP14			
DP15			

Table 2 shows by data partner (DP) which data and study diagnostics passed for the comparison of aflibercept and ranibizumab and table 3 shows by data partner (DP) which data and study diagnostics passed for the comparison of ranibizumab and bevacizumab. For all three study questions, every database that passed study diagnostics first passed data diagnostics. DP8 did not pass data diagnostics for the comparison of aflibercept to bevacizumab nor for the comparison of ranibizumab to bevacizumab and subsequently did not pass study diagnostics for these comparisons.

Table 2. Data and Study Diagnostic Results by Data Partner for the comparison of aflibercept and ranibizumab for the outcome end stage renal disease

	Passed Data Diagnostics	Ran Study Diagnostics	Passed Study Diagnostics
DP1			
DP2			
DP3			
DP4			
DP5			
DP6			
DP7			
DP8			
DP9			
DP10			
DP11			
DP12			
DP13			
DP14			
DP15			

Table 3. Data and Study Diagnostic Results by Data Partner for the comparison of ranibizumab and bevacizumab for the outcome end stage renal disease

	Passed Data Diagnostics	Ran Study Diagnostics	Passed Study Diagnostics
DP1			
DP2			
DP3			
DP4			
DP5			
DP6			
DP7			
DP8			
DP9			
DP10			
DP11			
DP12			
DP13			
DP14			
DP15			

Conclusion

Using the OHDSI SOS challenge as a pilot, Data Diagnostics was able to consistently and accurately identify databases that contain the necessary elements for given clinical questions with the potential to generate evidence. However, additional study diagnostics are needed to further determine if that evidence generated is reliable.

References

1. SOS Challenge – OHDSI [Internet]. [cited 2023 Jun 8]. Available from: https://www.ohdsi.org/sos-challenge/

2. OMOP CDM Database Diagnostics Utility [Internet]. [cited 2023 Jun 8]. Available from: https://ohdsi.github.io/DbDiagnostics/index.html

3. AntiVegfKidneyFailure/Protocol.Rmd at main · ohdsi-studies/AntiVegfKidneyFailure · GitHub [Internet]. [cited 2023 Jun 8]. Available from: https://github.com/ohdsi-studies/AntiVegfKidneyFailure/blob/main/Documents/Protocol.Rmd