

Name:	Christopher Knoll
Affiliation:	Janssen Research and Development
Email:	cknoll1@its.jnj.com
Presentation type (s):	Software Demonstration

Cohort Characterization in ATLAS with Standardized Feature Generation

Christopher Knoll, BCS^{1,2}, Ajit A. Londhe, MPH^{1,2}, Anthony G. Sena^{1,2}, Patrick B. Ryan, PhD^{1,2}

¹Janssen Research and Development, Titusville, NJ, ²Observational Health Data Sciences and Informatics, New York, NY

Abstract

Cohort characterization is recommended as a diagnostic step to explore both the baseline and post-index characteristics of patients in a study cohort. Demographics such as age, gender, race, and ethnicity, along with medical history and healthcare utilization can be informative about cohort's composition. Additionally, the dissemination of this characterization is recommended as a best practice for research papers by the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) methodology (1).

When designing an observational study, understanding the prevalence of attributes in each of the study's cohorts can help in determining if the cohorts, as constructed, should be leveraged in answering the study question. Well-designed studies comparing two or more cohorts contrast these cohort characterizations both before and after confounding adjustment through metrics like the standardized mean difference to evaluate the comparability of the study cohorts (2). Cohort characterization can also help in training patient level prediction models.

To facilitate the evaluation of a cohort's characteristics, standardized features based on patient exposures can be generated on a selection of baseline covariates. With the high volume of potential features often found in observational data, reporting tools are needed to enable easy aggregation and assessment of cohort characteristics.

Introduction

Atlas is an OHDSI tool that provides multiple capabilities to enable standardized scientific research and analysis (3). One feature that is used in cohort studies is the ability to define cohorts in a standardized way. However, exposing and exploring the cohort characteristics is currently limited, as there is no user interface to navigate and explore the prevalence of drug exposures, diagnoses, procedures, observations, and measurements. This software demonstration will show how Atlas would expose cohort characterization functionality to allow study designers to explore their cohorts of interest prior to executing their study.

Software Demonstration Topics

- Review the process for defining a cohort using Atlas cohort definition criteria.
- Describe the process for generating a "large scale" cohort characterization data set. Current capabilities consist of generating characterization information across domains of interest (demographics, conditions, drugs, procedures, measurements, and observations) for the defined cohort.
- Utilizing case study cohorts, show how the user can contextualize the prevalence of features through the hierarchical navigation of the OMOP Vocabulary.
- Create and export study tables based on user selected features and relative time windows.

Sample Screenshots

Column visibility Copy CSV Show 15 entries Filter:

Showing 1 to 9 of 9 entries Previous 1 Next

Name	Analysis Name	Domain	Time Window	Count	% of cohort
Persons by gender: gender = MALE	Persons by gender	Gender		1081	100.00
number of distinct persons	Person count	Person		1081	100.00
Age group: 35-39	Age group	Person		1	0.10
Age group: 40-44	Age group	Person		4	0.40
Age group: 45-49	Age group	Person		22	2.10
Age group: 50-54	Age group	Person		98	9.10
Age group: 55-59	Age group	Person		295	27.30
Age group: 60-64	Age group	Person		585	54.20
Age group: 65-69	Age group	Person		76	7.10

Figure 1. Prevalence report, for Demographics, Drug, Condition, Procedure and Measurement

Column visibility Copy CSV Show 15 entries Filter:

Showing 1 to 13 of 13 entries Previous 1 Next

Covariate Name	Domain	Count Value	Avg	Std Dev	Min	P10	P25	Median	P75	P90	Max	Z Score
Observation time prior to cohort start	Observation period	1355	1853.22	1324.69	365	508	757	1424	2727	3955	6151	1.78
Observation time after cohort start	Observation period	1355	387.46	329.30	0	56	136	293	553	832	2055	-1.89
Number of visits observed in 365d on or prior to cohort index	Visit	1355	47.11	28.74	0	18	27	40	61	85	209	2.13

Figure 2. Distribution report for continuous variables

References

1. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Journal of clinical epidemiology*. 2008;61(4):344-9.
2. Austin PC. Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research. *Communications in Statistics - Simulation and Computation*. 2009;38(6):1228-34.
3. OHDSI. OHDSI/Atlas. In: Informatics OHDSa, editor. *GitHub*2017.