#### A tool for empirically identifying and reviewing candidate comparators for pharmacoepidemiological studies

Justin Bohn, Jamie P. Gilbert, Christopher Knoll, David M. Kern, Patrick B. Ryan

## Background

The active comparator new user (ACNU) cohort design has emerged as a best practice for the estimation of drug effects from observational data[1]. However, despite its advantages, this design requires the selection and evaluation of comparators for appropriateness, a process which can be challenging. We developed a novel empirical approach to rank candidate comparators in terms of their similarity to a target drug in high-dimensional covariate space[2] and here demonstrate an interactive web application to explore candidate comparators ranked by our method.

# Methods

This first step in the similarity calculation pipeline is generating cohorts for comparison. We defined new user cohorts for each RxNorm ingredient and for each Anatomical Therapeutic Chemical (ATC) level 3 class across five administrative claims and two electronic health records databases. We then extracted aggregated pre-treatment covariate data for each cohort using five clinically oriented covariate domains: *demographics* (age and sex), *presentation* (conditions occurring  $\leq$  30 days before treatment), *medical history* (conditions occurring > 30 days before treatment), *prior medications* (medications begun > 30 days before treatment), and *visit context* (inpatient and emergency visits in the 30 days before treatment). In each database, we formed all pairs of cohorts with  $\geq$  1,000 patients and computed a scalar similarity score defined as the average of cosine similarities computed within each covariate domain for each pair. We then developed a web application[3] which allows users to input a target drug or drug class and receive a list of candidate comparators ranked by the overall similarity score. Users wishing to implement this method in their own data can do so using code available on GitHub[4].

# Results

In this demonstration, we describe several potential use cases for this tool, exploring its utility for a diverse set of example drugs. We show how users can adapt their queries to focus on specific databases or to require certain hierarchical relationships between target and comparator. Throughout, we show that the tool tends to produce lists of top-ranked comparators that broadly align with subject matter knowledge.

For example, **Figure 1**, **left** shows a user requesting rankings on RxNorm ingredient-level comparators for the multiple sclerosis treatment ocrelizumab using all available data sources and requiring comparator presence in at least two sources. The returned results (**Figure 1**, **right**) indicate a similar treatment for multiple sclerosis, natalizumab, as the most similar comparator, with data source-specific information available in a drop-down. By selecting "Explore Comparison", the user can then inspect the individual components of the overall similarity score (**Figure 2**, **top**) and compare the distribution of covariates in the target and comparator cohorts interactively (**Figure 2**, **bottom** and **Figure 3**).

#### Conclusion

This software demonstration will present a novel utility with which users can obtain empirical comparator recommendations. While the method underlying this tool produces recommendations that broadly align with subject matter knowledge and existing drug hierarchies, the utility of its recommendations may be limited in some cases. For example, recommendations for drugs that are used differently depending on dose, form, or route of administration, or which have multiple unrelated indications, or are used in combination therapies, may be less coherent than for drugs with more uniform characteristics. Nonetheless, this tool serves as a useful aid to investigators and could ultimately enable the automated generation of ACNU-derived evidence, a process that has previously been limited to self-controlled designs.

### **Figures**

Comparator Selection Explorer							
anssen Research & Development							
Recommend Comparators About							
Settings	Comparator listing						
Select target exposure:	Search						
RxNorm - ocrelizumab	Comparator				In ATC Class with Target		
Select comparator types:	Type Name	Overall Rank ↑	Average Similarity Score	Number of Databases	At Level 3	At Level 4	
RxNorm Ingredients							
Select data sources:	▼ RxNorm Ingredient natalizumab	1	0.980	4	Yes	Yes	Recommend Covariates for
IBM MDCD							Exclusion
IBM MDCR		Data Source	Sample Size	Cohort Simi	larity Score	Source	Specific Rank
V JMDC V Optum EHR	Explore Comparison	IBM CCAE	8,737	0.9	70		4 of 1,429
PharMetrics	Explore Comparison	IBM MDCD	2,034	0.9	77		1 of 1,171
PREMIER	Explore Comparison	Optum EHR	12,279	0.9	83		4 of 1,701
Minimum data sources with comparator presence:	Explore Comparison	PharMetrics	5,340	0.990		1 of 1,398	
	•						Recommend
Rank comparators on:	RxNorm Ingredient teriflunomide	2	0.979	3	Yes	Yes	Covariates
Average similarity score     Average source-specific rank							Exclusion

Figure 1: The user input panel (left) shows the selection of the target exposure, the desired comparator type(s), databases to draw similarity data from, and the methods for ranking across databases. The user has selected the RxNorm ingredient ocrelizumab, a popular treatment for multiple sclerosis, as the target. The main output pane (right) shows results based on the input. Note that the top comparator is natalizumab, a treatment for multiple sclerosis similar to ocrelizumab. Information is provided on database-specific similarity scores and ranks, as well as comparator sample size and relatedness in the ATC hierarchy. The user can select "Explore Comparison" to view the covariate profiles of the target and comparator. An option to "Recommend Covariates for Exclusion" allows the user to identify features that are very highly correlated to either the target or comparator and thus may inhibit propensity score model fitting.

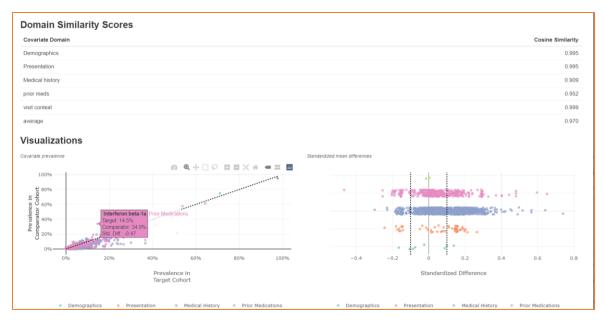


Figure 2: Details of the "Explore Comparison" window, showing the decomposition of the overall similarity score into cosine similarity values in five covariate domains. Figures interactively depict the prevalence of covariates in the target and comparator cohorts, as well as their standardized differences.

Demographics Presentation	Medical history Prior medicati	Prior medications Visit context Index date				
<b>ledical history</b> Ine covariate per condition observed and the covariate of the covariates (39.9%) have a		than 0.1				
Search						
Covariate	RxNorm - ocrelizumab (n = 7,678	) RxNorm - natalizumab (n = 8,737)	Std. Diff.			
Fatigue	36.6%	7.8%	0.74			
Vitamin D deficiency	51.3%	21.9%	0.64			
Anxiety disorder	26.6%	6.4%	0.57			
Paresthesia	30.0%	8.6%	0.57			
Anesthesia of skin	25.3%	6.9%	0.52			
Imaging of brain abnormal	20.9%	4.8%	0.50			
Asthenia	24.0%	6.8%	0.49			
Displacement of cervical intervertebral disc	17.6%	3.8%	0.46			
Gastroesophageal reflux disease without esophagitis	17.1%	3.7%	0.45			
Major depression, single episode	21.5%	7.1%	0.42			

Figure 3: Details of the "Explore Comparison" window show one of the five domain-specific covariate distribution tables. Users are informed of the number of covariates in each domain that are "in balance" (i.e., with absolute standardized difference < 0.1) and can search through the table for relevant features or re-sort the table to identify the most prevalent features in either the target or comparator cohort, or the largest differences between the two.

## **References/Citations**

- Lund JL, Richardson DB, Stürmer T. The Active Comparator, New User Study Design in Pharmacoepidemiology: Historical Foundations and Contemporary Application. *Curr Epidemiol Rep* 2015;**2**:221–8. doi:10.1007/s40471-015-0053-5
- 2 Justin Bohn, James P. Gilbert, Christopher Knoll, *et al.* Large-scale empirical identification of candidate comparators for pharmacoepidemiological studies. *medRxiv* 2023;:2023.02.14.23285755. doi:10.1101/2023.02.14.23285755
- 3 *Comparator Selection Explorer*. https://data.ohdsi.org/ComparatorSelectionExplorer/ (accessed 6 Jun 2023).
- 4 *ComparatorSelectionExplorer*. Observational Health Data Sciences and Informatics 2022. https://github.com/OHDSI/ComparatorSelectionExplorer (accessed 2 Jun 2023).