

Background

- Randomized controlled trials (RCTs) are often cited as the “gold standard” for evidence in medical research, but can have limited generalizability
- Electronic healthcare records (EHRs) and aggregation of such data in OMOP CDM format provides a potential resource for examining generalizability
- **Objective: To assess the homogeneity within and between a study (RCT) cohort and its potentially eligible target recruitment cohort**
- Case Study: A single center Phase II trial comparing rifaximin against placebo to determine which one reduces the amount of fibrosis in liver transplant patients with recurrent hepatitis C virus (HCV), occurring between Mar 2012 to Dec 2016

Methods

Construct cohorts

Select adult liver transplant patients with HCV

Study Cohort	43
Initial Target Cohort	136



Exclusion criteria applied to initial target*

Human immunodeficiency virus (HIV)**	5
Hepatitis B virus (HBV)**	3
Participation in another trial**	10
Females of childbearing age**	1
Tuberculosis**	0
Clostridium difficile within past 30 days	2

*Patients can qualify for multiple criteria

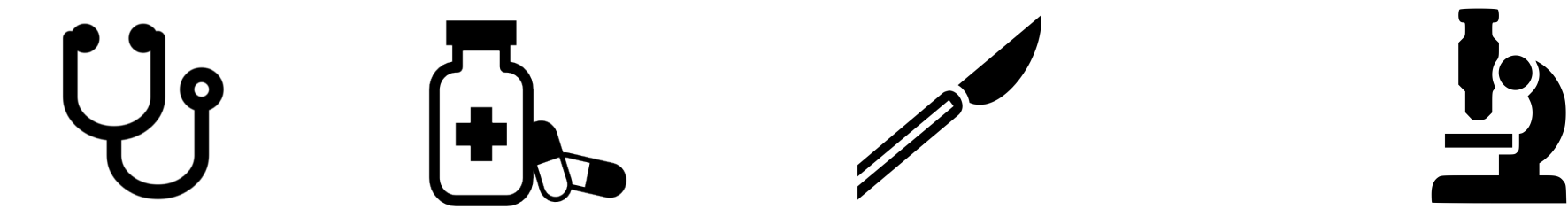
**Within past year of index date

Final cohort counts

Study Cohort	43
Final Target Cohort	116

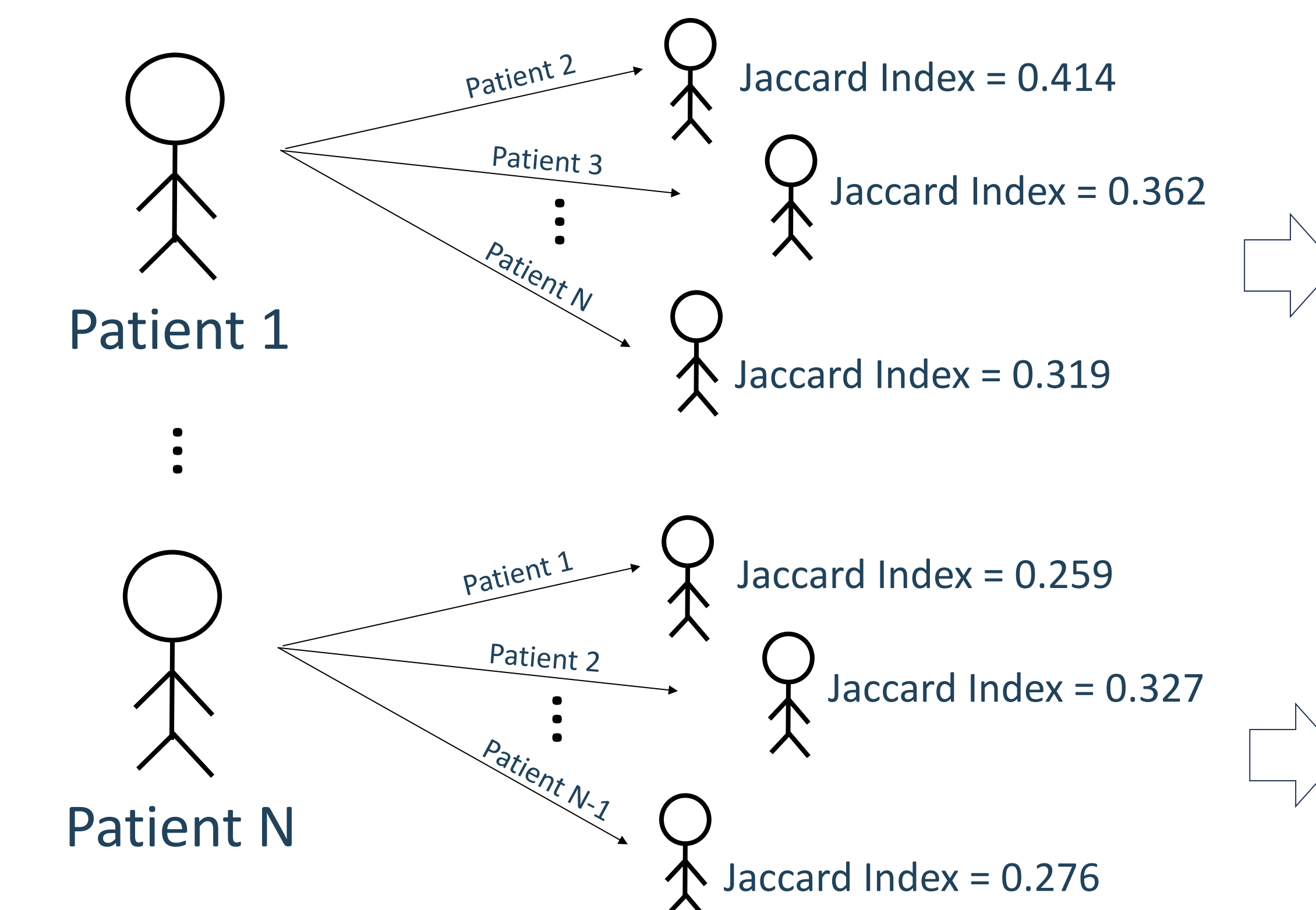
Data Setup

- For each cohort, extract all the following that occurred within one year of index:



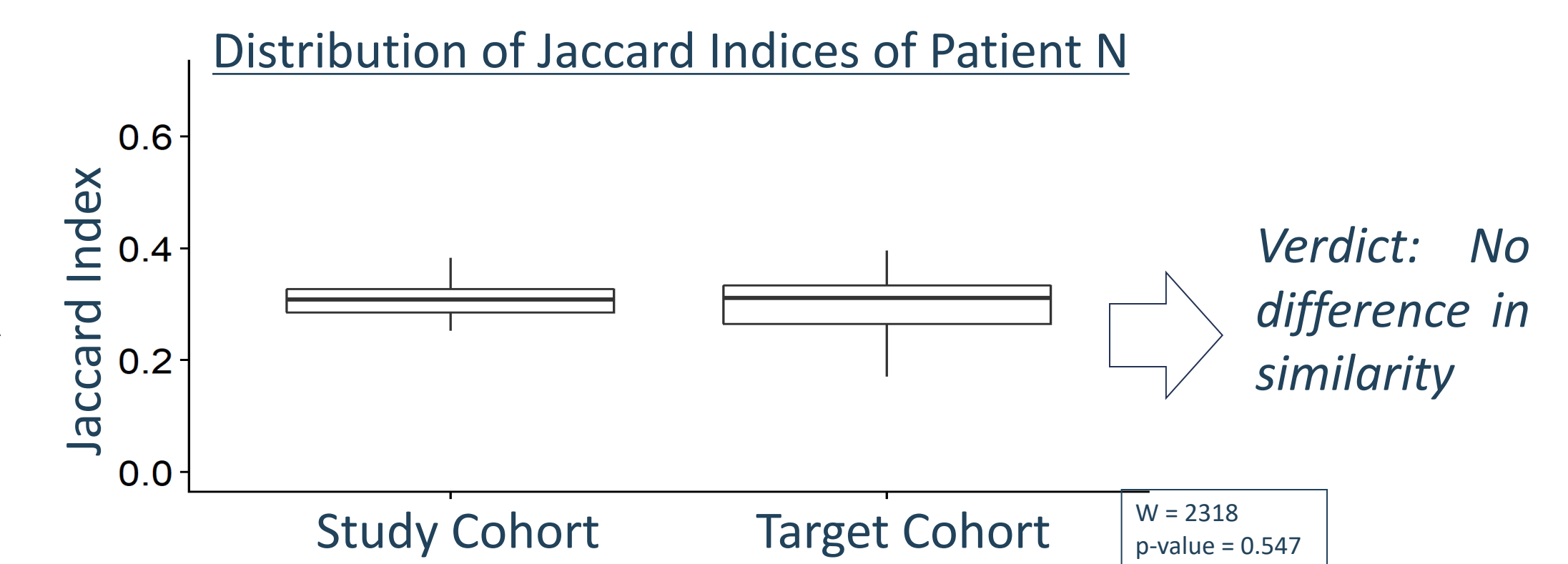
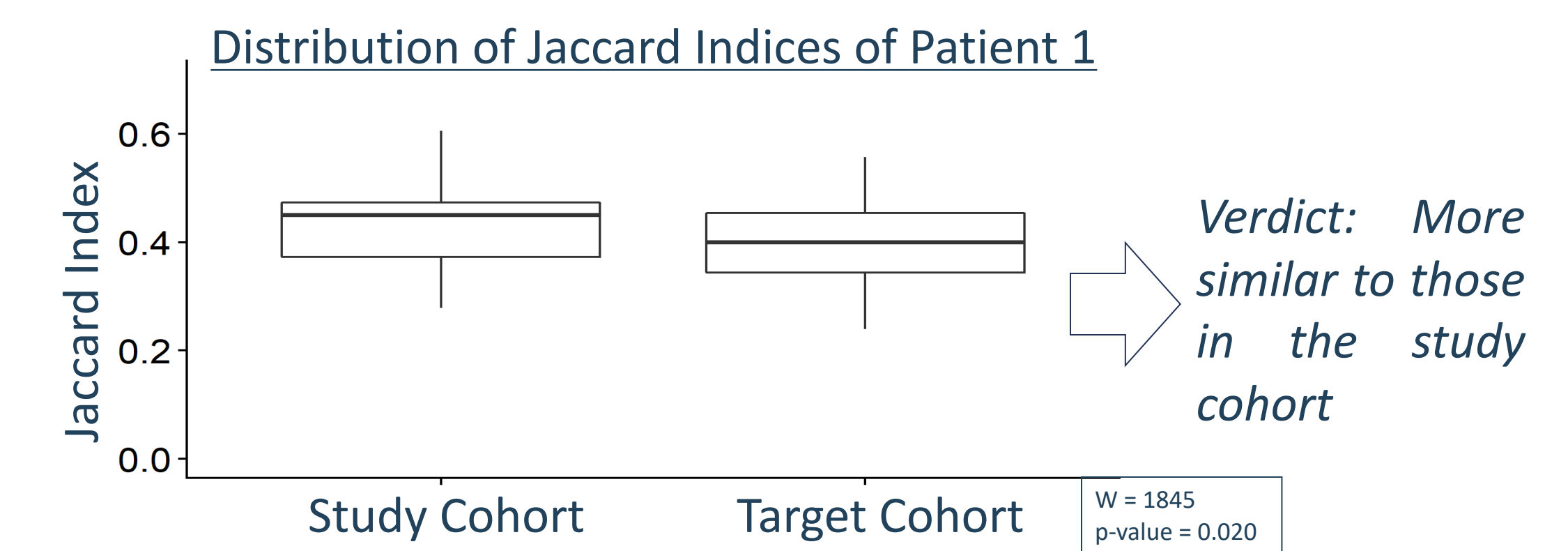
Conditions Drugs Procedures Measurements

- Then for each patient, calculate their Jaccard index relative to all other patients



Analysis

- **To investigate differences:** Explore standardized differences for baseline characteristics
- **To determine homogeneity:** Use a Mann-Whitney test on the distributions of the Jaccard indices per patient, stratified by study vs target, and aggregate results



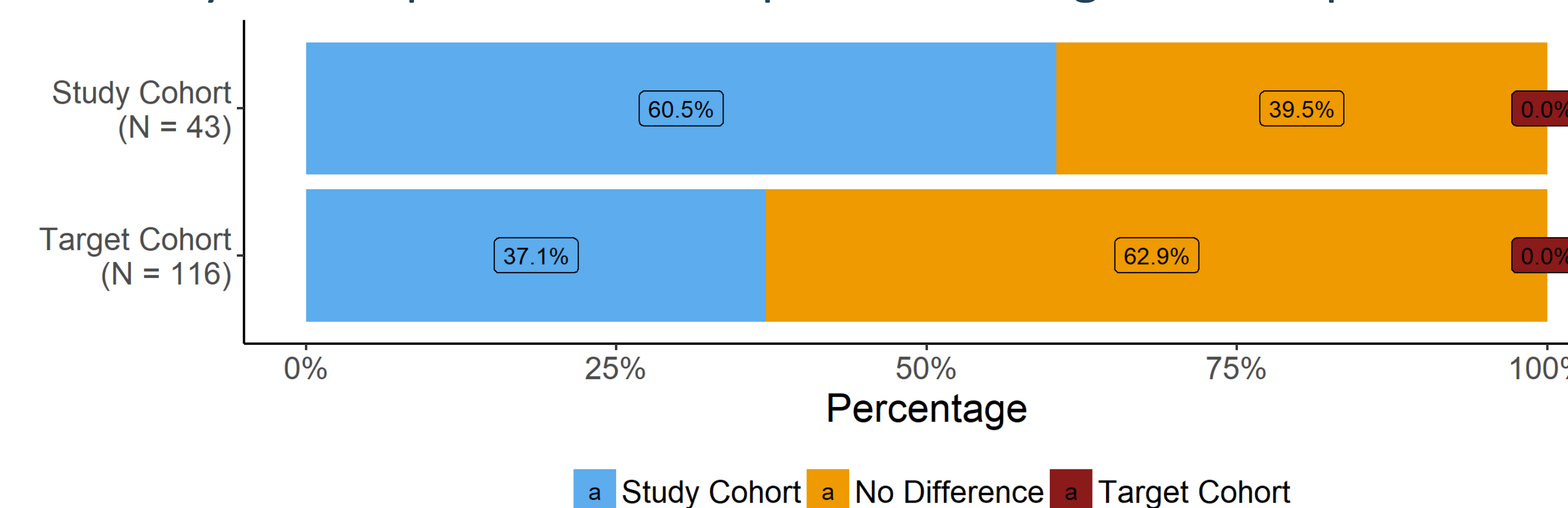
Results

- There were minimal differences found in baseline statistics between the study cohort and the target cohort for demographics and counts of unique concepts

	All N = 159 N (%) or mean ± SD	Study N = 43 N (%) or mean ± SD	Target N = 116 N (%) or mean ± SD	Standardized Difference Diff (95% CI)
Demographics				
Age	60.58 ± 7.36	60.93 ± 8.51	60.45 ± 6.91	0.06 (-0.29, 0.41)
Sex				
Male	116 (72.96)	31 (72.09)	85 (73.28)	0.03 (-0.32, 0.38)
Female	43 (27.04)	12 (27.91)	31 (26.72)	
Race				
White	93 (58.49)	29 (67.44)	64 (55.17)	0.25 (-0.10, 0.61)
Non-white or Other*	66 (41.51)	14 (32.56)	52 (44.83)	
Ethnicity				
Not Hispanic or Latino	88 (55.35)	24 (55.81)	64 (55.17)	0.01 (-0.34, 0.36)
Hispanic or Latino or Other*	71 (44.65)	19 (44.19)	52 (44.83)	
Concepts**				
Unique concepts	137.84 ± 60.22	136.02 ± 48.63	138.51 ± 64.17	0.04 (-0.31, 0.39)
Unique condition concepts	44.64 ± 22.67	44.47 ± 16.63	44.71 ± 24.59	0.01 (-0.34, 0.36)
Unique drug concepts	45.81 ± 23.83	44.37 ± 21.55	46.34 ± 24.69	0.09 (-0.27, 0.44)
Unique procedure concepts	47.39 ± 19.74	47.19 ± 17.38	47.47 ± 20.62	0.02 (-0.34, 0.37)
Unique measurement concepts	159.33 ± 51.20	158.51 ± 41.39	159.63 ± 54.55	0.02 (-0.33, 0.37)

*“Other” refers to either no matching concept, missing, or unknown
**All concepts were extracted within the past year of index date

- Most patients in the study cohort and almost 40% of patients in the target cohort were found to be more similar to other study cohort patients in comparison to target cohort patients



- There do exist a sample of a few differences between cohorts

Domain	Study N = 43 N (%)	Target N = 116 N (%)	Standardized Difference Diff (95% CI)
Midodrine	0 (0.00)	18 (15.52)	0.61 (0.25, 0.96)
Natriuretic peptide B [Mass/volume] in Serum or Plasma	0 (0.00)	18 (15.52)	0.61 (0.25, 0.96)
Transplantation of Liver, Allogeneic, Open Approach	0 (0.00)	15 (12.93)	0.55 (0.19, 0.90)
Hepatic encephalopathy	30 (69.77)	52 (44.83)	0.52 (0.17, 0.88)

Conclusions

- For this case study, there is fairly reasonable homogeneity between study cohort patients and target cohort patients
- Although the majority of the study cohort patients are similar to others, it is not a completely homogenous group that may lend itself to a prominent selection bias
- For the target cohort, the patients tend to be more heterogeneous relative the study cohort, but again demonstrate that minimal selection may exist
- This analysis can provide further guidance on evaluating study composition and interpretation