

PP03 Evaluating the Use of Real-World Data (RWD) as a Synthetic Control Group: A Statistical Approach

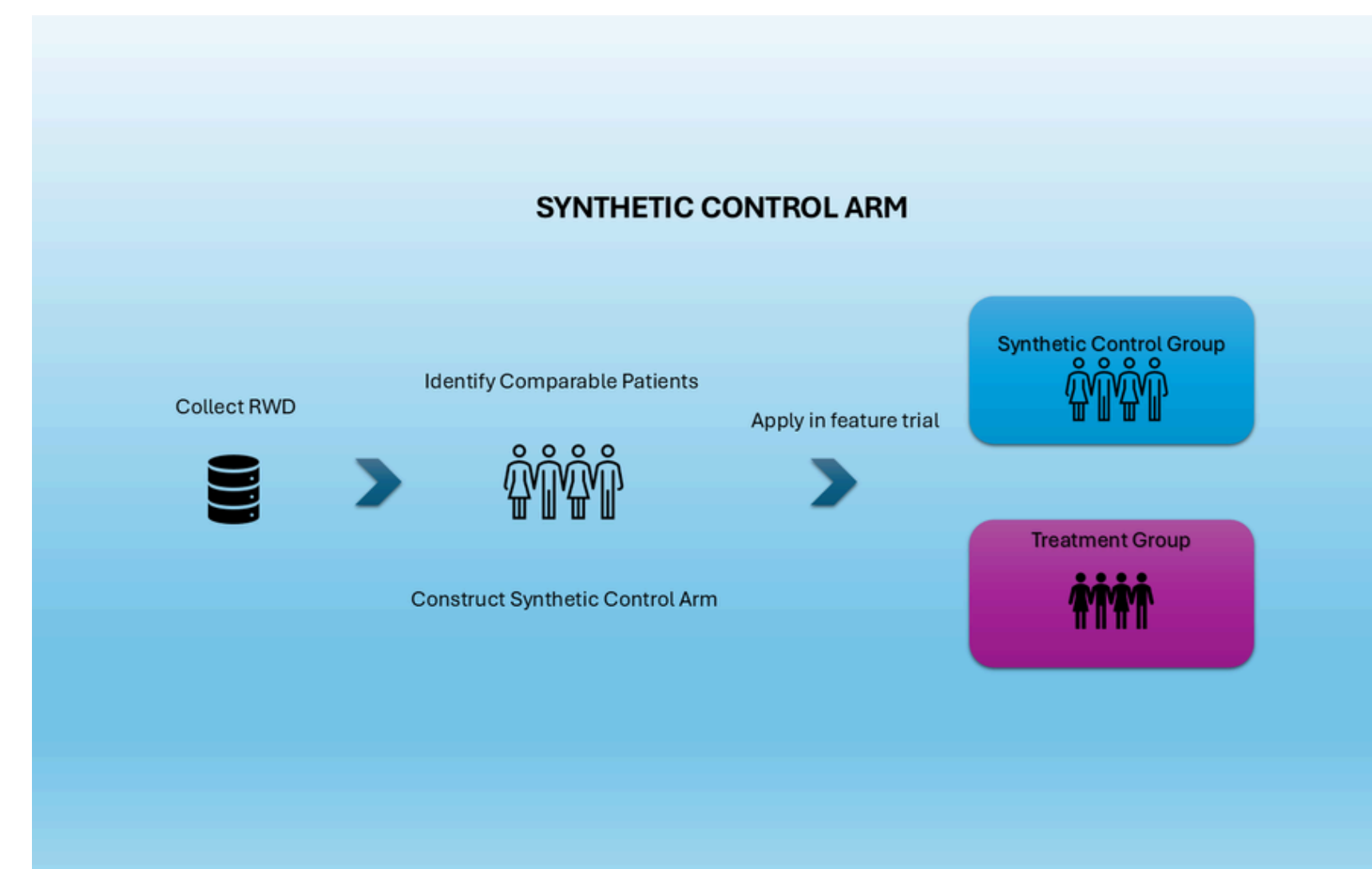


Authors: Economou Vasiliki|Vasiliki.Economou@pfizer.com, Kounelis Dimitrios|Dimitrios.Kounelis@pfizer.com, Kountouris Nikolaos|Nikolaos.Kountouris@pfizer.com

Introduction

Real World Data (RWD) are defined as the collection of patient health status data from sources like patient claims data and registries. A synthetic control arm (SCA) is a control group created from existing data, such as RWD, rather than by recruiting new participants in a clinical trial. It is used as a means of comparison with the treatment group, especially when it is costly or difficult to recruit a traditional control group, as in the case of research on rare diseases.

Fabry's disease is a rare, hereditary lysosomal storage disorder caused by a deficiency of the enzyme alpha-galactosidase A. Real-world data (RWD) analysis is essential for understanding rare diseases, but data scarcity poses challenges. To address this, we developed a synthetic RWD simulation dataset tailored to Fabry's disease, enabling the demonstration of practical RWD analysis methods.



Methodology

1. Definition of SCA

2. Balancing Groups PSM OR IPSW

3. Statistical Validation of Balance

4. Outcome Analysis

5. Sensitivity Analysis

1. Synthetic Data Exploration

- The study was based on bibliography and synthetic data made with R libraries.
- Population: 5,000 synthetic patients
- Treatment allocation: 4,477 on treatment, 523 on placebo
- Variables included:
 - Demographics: age, gender
 - Clinical symptoms: diffuse pain, skin lesions, gastrointestinal problems, vision problems, renal involvement,
 - Complications: cardiological complications, neurological complications, renal complications
 - Treatment-related: treatment response, treatment

2. Balancing Groups PSM-IPSW

PSM Approach

Nearest Neighbor Matching
Caliper: 0.1 SD
1:1 Matching Ratio

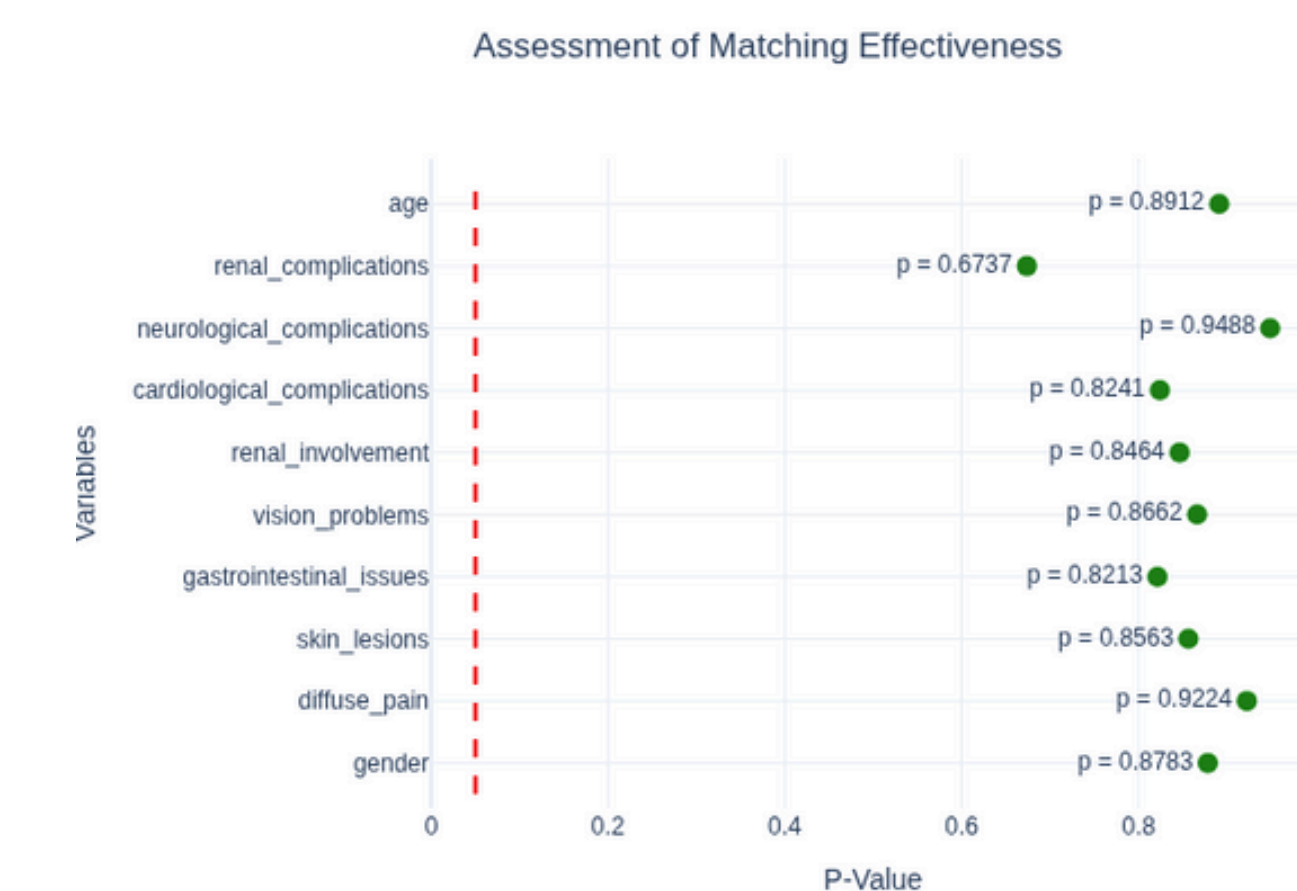
IPSW Approach

Inverse Probability Weighting
Weight=1/PS

IPSW>PSM

1. Better covariate balance
2. Preserves full sample
3. Ideal for rare diseases

3. Statistical Validation of Balance



*To confirm that the balancing method applied to the two groups (IPSW) was effective, we expect all p-values to be greater than 0.05.

4. Outcome Analysis

4.1 Association and Multicollinearity Analysis

Variable	P-Value	VIF
Treatment	<0.0001	1.001721
Skin_lesions	<0.0001	1.03467
GI issues	0.0016	1.023369
Renal_involvement	0.0057	1.062689

*The step before logistic Regression is to check the association between the dependent and independent variable and the multicollinearity between the variables.

The rest of the variables had $VIF < 2$ but the correlation with the dependent variable is statistically insignificant.

4.2 Logistic Regression

Observations	Independent variables	Pseudo-R ² (Cox & Snell)	Nagelkerke R ²
5,000	11	0.3348	0.2463

Hosmer-Lemeshow: $\chi^2 = 13.448$, $p = 0.097 \rightarrow$ Adequate fit ($p > 0.05$)

Significant Variables	OR	β	P-Value
Treatment	6.66	1.89	<0.001
Skin_lesions	1.32	0.27	<0.001
GI issues	1.24	0.21	<0.001
Renal_involvement	0.79	-0.23	0.001

Treatment OR ≈ 6.66 , patients who received treatment were approximately 6.7 times more likely to respond positively.

Non-significant predictors ($p > 0.05$)

- Age
- Gender
- Diffuse pain
- Vision problems
- Cardiac complications
- Neurological complications
- Renal complications

5. Sensitivity Analysis

SENSITIVITY ANALYSIS: E-VALUE ASSESSMENT

12.80
E-VALUE

INTERPRETATION:

Unmeasured confounder would need $RR \geq 12.80$ association with BOTH treatment AND outcome to explain away observed effect.

✓ ALREADY CONTROLLED:

Age, Gender, Renal/Cardiac/Neurological complications, Disease severity, Comorbidities

To evaluate the robustness of the treatment effect to potential unmeasured confounding, we calculated the E-value.

The odds ratio (OR) for the effect of treatment on response was estimated at 6.66, indicating a strong and statistically significant association.

Conclusion

This study demonstrates the feasibility and methodological rigor of using real-world data to construct synthetic control groups in rare disease research. By leveraging advanced weighting techniques such as IPSW, researchers can achieve superior covariate balance, preserve sample size, and enhance computational efficiency compared to traditional matching methods.

The robust treatment effect estimates, supported by comprehensive sensitivity analyses, underscore the potential of this framework to generate clinically meaningful and regulatory-acceptable evidence. These findings highlight synthetic control methodologies as a powerful tool for overcoming the inherent challenges of rare disease studies, paving the way for more efficient, ethical, and generalizable clinical research.