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

Vasiliki Economou, Dimitrios Kounelis, Nikolaos Kountouris

1.Abstract

Real World Data (RWD) are defined as the collection of patient health status data from sources like patient claims data and registries. The use of RWD as a Synthetic Control Arm (SCA) is an innovative approach that assists in evaluating therapeutic interventions with data derived from a real clinical situation. The SCA method is particularly useful when dealing with rare or hard to detect conditions, where discovering numerous participants can prove to be impossible and time consuming. Regulators are increasingly recognizing the value of SCA, with recent guidance frameworks facilitating its integration into some processes. We address data heterogeneity and selection bias by applying advanced statistical methods like propensity score matching and prognostic modelling, enabling the use of RWD-based external controls in rare disease research. Appropriate use of RWD-based external controls can assist and accelerate therapeutic developments for patients with rare diseases waiting for innovative therapies.

2.Introduction

Real World Data (RWD) are defined as the collection of patient health status data from sources like patient claims data and registries. We can collect RWD data from electronic health records (EHRs), insurance billing and claims, disease or medication/device registries and Patient-reported outcomes. Using these sources of data, therapies can be evaluated under real-world conditions in a broader population and at a much lower cost.

A synthetic control arm (SCA) is a control group created from existing data, such as RWD, rather than by recruiting new participants. 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. More specifically, traditional clinical trials require significant resources for recruiting, managing, and monitoring control groups. Synthetic control groups, by leveraging existing data, drastically reduce these costs, making clinical trials more economically viable. In addition, the process of finding, recruiting, and managing participants in a control group can be time-consuming. SCAs simplify this process, potentially leading to faster trial completion and faster access to new treatments. In some cases, it may be unethical to deny patients a treatment that may benefit them by placing them in a control group. SCAs allow the use of synthetic control arms, eliminating the need for some patients to receive a placebo or standard treatment, while still allowing researchers to evaluate the effectiveness of the new treatment. By leveraging multiple data sources, SCAs can mitigate potential biases that may arise from traditional control group selection, such as those related to patient demographics, comorbidities, or geographic location.

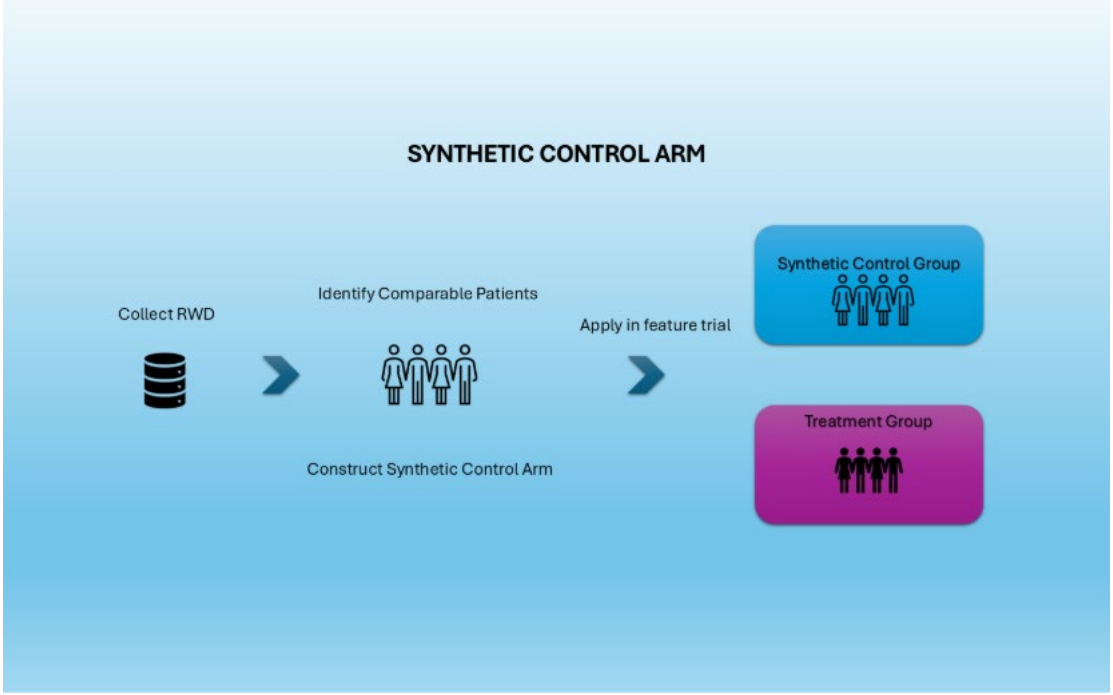


Figure 1 Synthetic Control Arm visual description

3.Methods

To ensure validity and minimize bias, specific standards are applied to non-randomized data analysis. This is vital to establish a synthetic control group (SCA) from Real-World Data (RWD) for comparison with an intervention group that received the new treatment under investigation.
Initially, when forming the control group, the inclusion and exclusion criteria should be considered, which are identical to those of the clinical trial. We want RWD data, such as electronic medical records and patient registries, to resemble the target population for treatment.

To avoid bias, we will apply methods that aim to balance the treatment and control groups based on observed baseline characteristics. One common approach is Propensity Score Matching (PSM), which estimates the probability that a patient receives the treatment under study, using logistic regression with covariates such as age, gender, comorbidities, functional indicators, laboratory results, and previous treatments. This allows for the creation of a matched control group that is comparable to the treatment group. Inverse Probability of Sampling Weighted (IPSW) uses weights based on propensity scores to create a synthetic population where the distribution of baseline covariates is independent of treatment assignment. Each subject's weight equals the inverse of the probability of receiving their actual treatment, effectively up-weighting under-represented populations and down-weighting over-represented ones. The key advantages of IPSW include its ability to retain all observations in the analysis and its flexibility in estimating different causal parameters. IPSW can estimate the average treatment effect (ATE) in the entire population, the average treatment effect in the treated (ATT), or the average treatment effect in the control (ATC), depending on the weighting scheme employed.

Characteristic	PSM	IPSW
Approach	Matches patients with similar propensity scores	Weights patients by inverse treatment probability
Sample Size	Reduced (matched pairs only)	Preserved (all patients retained)
Implementation	Direct analysis on matched cohort	Weighted analysis with all data
Causal Parameter	Average Treatment Effect in Treated (ATT)	ATE, ATT, or ATC
Best for	Large samples, clear matching criteria	Small samples, rare diseases

Figure 2 Comparison of PSM with IPSW

After balancing, we can also prove statistically that the grouping has been done correctly. This depends on the nature of the data: when continuous variables do not follow a normal distribution, nonparametric tests, such as the Wilcoxon Rank-Sum test, can be applied to compare their distributions between groups. For categorical variables, the Chi-square (X^2) test is used to examine whether the distribution of categories differs significantly between groups. The purpose of using these tests is to check whether, after balancing, statistically significant differences remain between the treatment group and the control group in their basic characteristics. Ideally, these tests should not show significant differences, supporting the claim that the groups are adequately balanced.

The effect of treatment is estimated using appropriate statistical models of the post-balancing groups, depending on the nature of the outcome. For time-to-event outcomes, Cox proportional hazards models are used, which allow the estimation of hazard ratios and can incorporate confounding variables or weights derived from IPTW or PSM. At the same time, the construction of Kaplan–Meier curves provide a nonparametric representation of survival per group, while the log-rank test is used to compare the curves. In the case of binary outcomes, such as the occurrence or non-occurrence of a clinical event, generalized linear models with logistic regression are applied, which attribute the probabilities of occurrence of the event per group and allow the calculation of odds ratios or risk ratios, depending on the frequency of the event and the study design. These models can incorporate IPTW weights or be applied directly to the sample created with PSM. For continuous outcomes, such as quality-of-life indices or biochemical parameters, linear regression models are applied to estimate differences in means between groups. In all the above cases, the incorporation of weights is critical to eliminate residual confounding as much as possible and maintain the validity of the causal estimation.

Sensitivity analysis is a critical step in studies with non-randomized data, as it allows for the evaluation of the stability of findings in the face of uncertainties, such as the existence of unobserved confounding factors or model assumptions. The application of such analyses concerns not only the statistical significance of the results, but mainly the causal interpretation and generalizability of the conclusions. A widely used technique is tipping-point analysis, which attempts to answer the question: how strong an unobserved confounding factor would have to be for the observed effect of the treatment to no longer be statistically or clinically significant? In other words, it identifies the “tipping point” beyond which the result will be reversed. This approach allows researchers and decision-makers to understand the level of uncertainty that may exist around the results, without necessarily knowing the exact nature of the confounding factor. Beyond that, Quantitative Bias Analysis (QBA) offers a more analytical tool, allowing assumptions to be made about the impact of an unobserved factor on the outcome and the probability of exposure. Through this process, it is estimated how much and in what direction the result is altered due to this bias. QBA can be implemented through methods such as calculating the E-value, which represents the minimum size of the relationship between an unobserved factor and both the outcome and the treatment that would be able to fully explain the observed result.

Figure 3 Flow chart for constructing the SCA

4. CASE STUDY – EXAMPLE

4.1 Example Introduction

To demonstrate the practical application of real-world data (RWD) analysis methods, we developed a synthetic RWD simulation dataset tailored to Fabry's disease, a rare and hereditary lysosomal storage disorder caused by a deficiency of the enzyme alpha-galactosidase A. The disease manifests with a heterogeneous clinical picture affecting multiple organs and systems, such as the renal, cardiovascular, and nervous systems.

The primary objective of this study was to compare the results between two groups of patients who received two different treatments. One group was the treatment group and the other the control group. Boths groups were created from historical or observational data sources. The target of the Synthetic Control Arm (SCA) is to operate as a "substitute control group" which can serve as an actual control group made up of existing patients.

The synthetic data generation was based on bibliography and was achieved using Rstudio (specifically the simstudy and sunthpop R libraries). The whole analysis was conducted in Python. The total population of the dataset was 5000 patients, of which 4477 patients belong to the treatment group and the rest 523 belong to the control group.

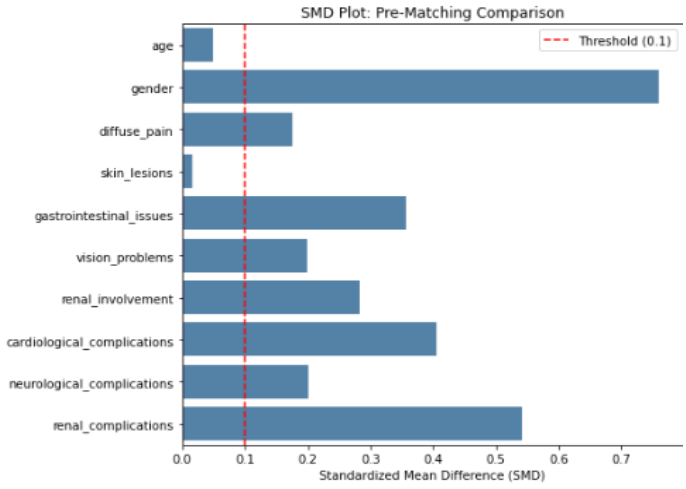
4.2 Balancing Groups

In this first phase of the analysis, the primary objective was to ensure the balance of covariates between the treatment group and the control group. To achieve this, two widely accepted methodologies were used: propensity score matching (PSM) and inverse probability of sampling weights (IPSW). Both approaches were applied using clinical and demographic variables derived from synthetic data specific to Fabry's disease. The selection of variables used for adjustment was based on a focused literature review of clinical studies and observational analyses related to Fabry's disease. These variables—age, sex, diffuse pain, skin lesions, gastrointestinal problems, vision problems, renal involvement, cardiological complications, neurological complications, renal complications, treatment response and treatment—have been previously reported as clinically important predictors of disease burden and treatment response. The inclusion and exclusion criteria were harmonized with those used in the clinical evaluation protocols of the new treatment, ensuring consistency and comparability between populations.

Figure 4 illustrates the distribution of standardized mean differences (SMDs) for the covariates before and after PSM and IPSW. A commonly accepted threshold of 0.1 was used to assess the adequacy of balance (indicated by the red dotted line in Figure 4 below).

Before matching, significant imbalance was observed in several covariates, exceeding the 0.1 threshold. After PSM, a significant improvement was achieved, although some variables remained marginally imbalanced. After IPSW, the overall balance of interactions improved further, with most variables falling below the 0.1 threshold, indicating the superior performance of IPSW in this simulation scenario.

The analysis of standardized mean differences (SMDs) shows that the IPSW method provided overall better balancing of confounding variables, as most variables had values below the acceptance threshold (0.1). The fact that the method IPSW had slightly better results is reflected in the previous section on the theoretical level of each method.



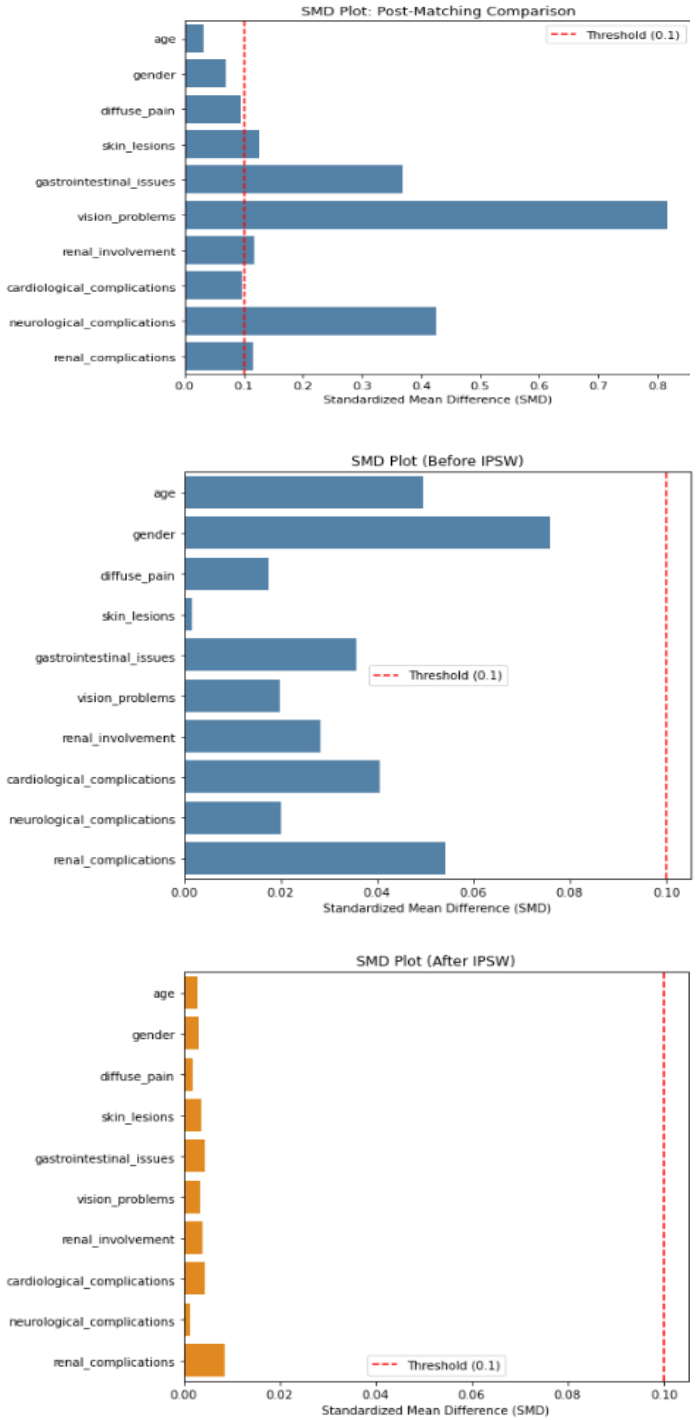


Figure 4 Comparison of Variable Balance: Before and After PSM and IPSW Application

4.3 Statistical Validation of Balance

To evaluate the effectiveness of the balancing procedure, we performed Chi-square tests for categorical baseline variables to assess differences between treatment and control groups. The p-values for all variables (gender, diffuse pain, skin lesions, gastrointestinal issues, vision problems, renal involvement, cardiological complications, neurological complications, and renal complications) were greater than 0.05, indicating no statistically significant differences post-balancing. This suggests that the groups were well balanced in terms of these observed baseline characteristics. The detailed results are presented below:

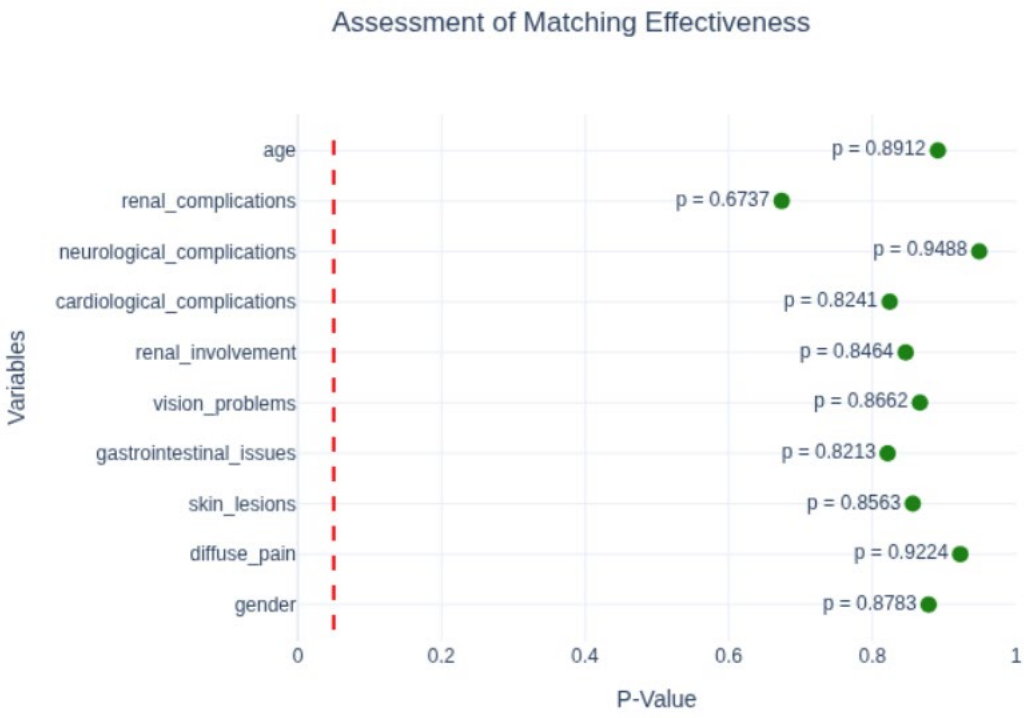


Figure 5 Assessment of Matching Effectiveness

4.4 Outcome Analysis

Next, we want to assess whether patients receiving treatment have better outcomes than those receiving a placebo. Given the binary nature of our outcome variables, logistic regression models were used for all analyses rather than time-to-event models such as Cox proportional hazards of regression. This approach treats each clinical manifestation as a binary endpoint. Logistic regression was applied to assess the effect of treatment on the probability of a positive response (treatment_response), while controlling for possible confounding variables.

Prior to applying regression analysis, a preliminary analysis was conducted to evaluate the correlations between the independent variables and the dependent variable (treatment_response). For the numerical variable age, an independent samples t-test was used, while for the binary variables, either the Chi-square test or Fisher's exact probability test was applied, depending on the number of observations per cell. The results showed that the treatment variable is statistically significantly related to the response to treatment ($p < 0.001$), while some other variables, such as renal_involvement, showed marginal statistical significance ($p \approx 0.029$).

In addition, the presence of multicollinearity among the independent variables was checked using the Variance Inflation Factor (VIF). All VIF values ranged between 1.00 and 1.17, indicating no multicollinearity. Therefore, it was deemed appropriate to retain all independent variables in the final logistic regression model.

Variable	VIF	p-value
const	20.560122	-
age	1.0016	0.8425
treatment	1.001721	**0.0**
gender	1.166016	0.643
diffuse_pain	1.069977	0.5185
skin_lesions	1.03467	**0.0**
gastrointestinal_issues	1.023369	**0.0016**
vision_problems	1.002369	0.8396
renal_involvement	1.062689	**0.0057**
cardiological_complications	1.086089	0.282
neurological_complications	1.001078	0.254
renal_complications	1.001976	0.159

Table 1 shows the result of multivariability and the second image shows the result of comparing the variables with the outcome.

The analysis was based on 5,000 observations and included 11 independent variables. The regression model demonstrated a satisfactory overall fit. The pseudo-R-squared (Cox & Snell) was 0.3348, indicating that approximately 33.5% of the variance in the outcome is explained by the independent variables. The Nagelkerke R^2 was 0.2463, suggesting a moderate fit of the model on a scale that can reach 1. Furthermore, the Hosmer–Lemeshow test (χ^2 = 13.448, p = 0.097) indicates the adequacy of the model fit to the data. Although a pseudo-R^2 of around 33% may seem modest compared to linear regression standards, it is considered satisfactory in rare disease analyses because logistic regression for rare events typically yields lower goodness-of-fit measures due to the limited variation in the outcome (King & Zeng, 2001).

The treatment variable was statistically significant ($p < 0.001$) with a coefficient of 1.8960 (95% CI: 1.808 – 1.984). This corresponds to an odds ratio of ≈ 6.66, suggesting that patients who received treatment were approximately 6.7 times more likely to respond positively than those who did not, considering all other variables.

4.4.1 Significant variables

Skin lesions: positively correlated with response (β = 0.2752, $p < 0.001$, OR ≈ 1.32)
Gastrointestinal issues: positively associated (β = 0.2132, $p < 0.001$, OR ≈ 1.24)
Renal involvement: negatively associated (β = -0.2348, p = 0.001, OR ≈ 0.79)
Treatment: positively correlated with response (β = 1.8960, $p<0.001$, OR ≈ 6.66)

4.4.2 Non-significant variables

The remaining variables, such as age, gender, diffuse pain, vision problems, cardiac or neurological complications, and renal complications, did not show a statistically significant correlation with response ($p > 0.05$).

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-0.8669	0.080	-10.800	0.000	-1.024	-0.710
treatment	1.8960	0.045	42.178	0.000	1.808	1.984
age	0.0011	0.001	0.859	0.390	-0.001	0.003
gender	-0.0152	0.052	-0.291	0.771	-0.118	0.087
diffuse_pain	-0.0452	0.046	-0.976	0.329	-0.136	0.046
skin_lesions	0.2752	0.054	5.144	0.000	0.170	0.380
gastrointestinal_issues	0.2132	0.060	3.538	0.000	0.095	0.331
vision_problems	-0.0235	0.071	-0.332	0.740	-0.162	0.115
renal_involvement	-0.2348	0.071	-3.319	0.001	-0.373	-0.096
cardiological_complications	-0.0741	0.075	-0.993	0.321	-0.220	0.072
neurological_complications	-0.1371	0.101	-1.353	0.176	-0.336	0.061
renal_complications	0.1787	0.104	1.722	0.085	-0.025	0.382

Figure 6 The results of the regression analysis

5. Sensitivity Analysis

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.

The corresponding E-value was 12.80, suggesting that an unmeasured confounder would need to be associated with both the treatment assignment and the outcome by a risk ratio of at least 12.80, above and beyond the measured covariates, to fully explain away the observed treatment effect.

This is a very high threshold, and it is unlikely that such a strong unmeasured confounder exists, particularly given that major prognostic variables (e.g., age, gender, renal involvement, and other complications) were already included in the model. Thus, the results appear robust to potentially unmeasured confounding.

In other words, the effect of treatment is unlikely to be an artifact of hidden bias unless there exists an extremely strong and unaccounted factor — one that is more strongly associated with both treatment and outcome than any measured variable in the model.

6.Discussion

Principal Findings and Methodological Framework

This analysis shows how a synthetic control group can be created using RWD. This approach demonstrates how electronic health records, disease registries and claims data can be systematically leveraged to create comparison groups that mirror the characteristics of treated populations while avoiding the recruitment challenges and ethical concerns associated with traditional control groups. The successful implementation of this framework in Fabry's disease research illustrates the broader applicability of synthetic control approaches across rare disease contexts, particularly for conditions where patient populations are geographically dispersed, disease progression is heterogeneous, and treatment options are limited.

Covariate Balancing and Methodological Superiority

The comparative evaluation of propensity score matching versus inverse probability of sampling weights revealed fundamental differences in their ability to address confounding in rare disease contexts. In the analysis it is demonstrated that while both methods can achieve meaningful covariate balance, IPSW consistently outperformed PSM across multiple critical dimensions including sample preservation, balance achievement, and computational efficiency. The superior performance of IPSW in retaining all available patients while achieving ideal covariate balance addresses a critical limitation in rare disease research where every patient represents valuable and often irreplaceable information.

The theoretical advantages of IPSW over PSM become particularly pronounced in rare disease applications where complex, multi-system involvement creates intricate confounding patterns that are difficult to address through individual patient matching. The ability of IPSW to simultaneously optimize balance across multiple covariates through differential weighting provides a more flexible and mathematically elegant solution to confounding adjustment than the discrete matching approach employed by PSM. Furthermore, the computational advantages of IPSW, including faster processing times and reduced implementation complexity, facilitate reproducibility and standardization across research teams and regulatory submissions, addressing practical concerns that often limit the adoption of sophisticated analytical methods in clinical research settings.

Balance Assessment and Quality Control

The rigorous assessment of covariate balance using standardized mean differences and complementary statistical tests demonstrates the critical importance of methodological quality control in synthetic control applications. This systematic approach to balance evaluation, incorporating both statistical metrics and visual assessment tools, provides a comprehensive framework for validating the success of confounding adjustment before proceeding to outcome analysis. The achievement of standardized mean differences below 0.1 across all covariates with IPSW, compared to residual imbalance with PSM, illustrates how seemingly minor methodological differences can have substantial implications for analytical validity and regulatory acceptance.

Treatment Effect Estimation and Clinical Significance

The outcome analysis component of the study demonstrates how properly balanced synthetic control groups can yield clinically meaningful and statistically robust treatment effect estimates. The observed odds ratio of 6.66 for treatment effectiveness represents a substantial clinical benefit that aligns with established understanding of therapeutic interventions in Fabry's disease, where enzyme replacement therapy and substrate reduction therapy can dramatically alter disease progression when initiated appropriately. The statistical significance of this finding, combined with narrow confidence intervals that exclude the null value, provides strong evidence for genuine treatment efficacy rather than methodological artifact.

The clinical interpretation of treatment effects must consider both statistical significance and clinical meaningfulness, particularly in rare disease contexts where effect sizes may be large, but confidence intervals wide due to small sample sizes. This case study demonstrates how synthetic control methodologies can achieve sufficient statistical power to detect clinically important effects while maintaining the analytical rigor necessary for regulatory acceptance. The identification of specific patient characteristics associated with treatment response, including the positive associations with early disease manifestations and negative associations with advanced organ involvement, provides clinically actionable insights that can inform treatment decision-making and patient counseling in clinical practice.

Sensitivity Analysis and Causal Inference

The comprehensive sensitivity analysis framework employed in this study addresses the fundamental limitation of observational research regarding potential unmeasured confounding. The E-value methodology provides a quantitative assessment of the robustness of treatment effect estimates to potential hidden bias, offering regulators and clinicians a metric for evaluating the credibility of causal claims derived from observational data. The exceptionally high E-value of 12.80 observed here indicates that an unmeasured confounder would need to be associated with both treatment assignment and clinical outcomes by a risk ratio exceeding any known clinical factor to explain away the observed treatment effect.

Clinical Practice Integration and Future Applications

The translation of these methodological advances into clinical practice requires consideration of both the technical implementation and the broader healthcare context in which rare disease treatment decisions are made. The evidence generated through synthetic control methods can inform clinical decision-making by providing real-world effectiveness data that complements efficacy findings from traditional clinical trials.

The scalability of synthetic control methodologies across multiple rare diseases and therapeutic contexts represents a significant opportunity for advancing rare disease research more broadly. The framework developed for Fabry's disease can be adapted to other lysosomal storage disorders, metabolic diseases, and genetic conditions where similar analytical challenges exist. The establishment of standardized protocols for synthetic control implementation, quality assessment, and regulatory submission could facilitate more efficient drug development processes and accelerate patient access to innovative therapies across the rare disease spectrum.

Limitations and Methodological Considerations

While this study demonstrates the potential of synthetic control methodologies in rare disease research, several limitations must be acknowledged that may influence the interpretation and generalizability of findings. The reliance on simulated data, although carefully designed to reflect realistic Fabry's disease characteristics, may not capture the full complexity of real-world clinical scenarios including data quality issues, missing information patterns, and temporal dynamics that could affect analytical performance. The cross-sectional analytical framework, while appropriate for methodological comparison purposes, may not fully address the progressive nature of rare diseases where treatment effects and confounding patterns evolve over time.

The external validity of findings beyond Fabry's disease requires careful consideration of disease-specific factors that may influence the relative performance of different synthetic control approaches. Conditions with different inheritance patterns, progression trajectories, or treatment mechanisms may present unique analytical challenges that could alter the optimal methodological choices. The successful implementation of synthetic control methods in clinical practice will require ongoing validation across multiple rare disease contexts, development of user-friendly analytical tools, and training of research teams in sophisticated causal inference techniques.

Future Research Directions and Innovation

The methodological framework established in this study provides a foundation for several important research directions that could further advance synthetic control applications in rare disease research. The integration of longitudinal analytical approaches that can better capture disease progression and time-varying treatment effects represents a critical next step for addressing the temporal complexity inherent in rare disease research.

The incorporation of biomarker data, genomic information, and other molecular-level measurements into synthetic control frameworks could provide more precise confounding control and enable earlier detection of treatment effects before clinical manifestations become apparent. The extension of synthetic control methods to network analyses that leverage information across related rare diseases or similar therapeutic mechanisms could expand effective sample sizes while respecting disease-specific characteristics. These methodological advances, combined with improvements in real-world data quality and availability, position synthetic control approaches as increasingly viable alternatives to traditional experimental designs for rare disease research.

Conclusions and Broader Impact

This comprehensive evaluation of synthetic control methodologies demonstrates both the technical feasibility and clinical value of sophisticated observational research approaches in rare disease contexts. The superior performance of IPSW over PSM, combined with robust sensitivity analysis frameworks and clinically meaningful treatment effect estimation, establishes a new standard for observational evidence generation that can support regulatory decisions and clinical practice improvements. The methodological rigor demonstrated here addresses historical limitations of observational research while providing practical solutions to the unique challenges faced in rare disease research.

The broader impact of this work extends beyond methodological advancement to encompass improvements in patient care, accelerated drug development, and more efficient healthcare resource allocation in rare disease contexts. By demonstrating how real-world data can be systematically leveraged to generate high-quality evidence, this study supports the evolution toward more flexible and responsive evidence generation paradigms that better serve the needs of rare disease communities. The establishment of comprehensive analytical frameworks for synthetic control implementation provides researchers, regulators, and clinicians with the tools necessary to extract meaningful insights from observational data while maintaining the analytical rigor essential for supporting causal claims and informing healthcare decisions.

References

1. Allison, P. (2019). Evaluating large-scale propensity score performance through real-world and synthetic data experiments. *International Journal of Epidemiology, 48*(6), 1834–1844. https://doi.org/10.1093/ije/dyz123
2. Bouttell, J., Craig, P., Lewsey, J., Robinson, M., & Popham, F. (2018). Synthetic control methodology as a tool for evaluating population-level health interventions. *Journal of Epidemiology & Community Health, 72*(8), 673–678. https://doi.org/10.1136/jech-2017-210106
3. Goldsack, J. (2019, February 5). Synthetic control arms can save time and money in clinical trials. *STAT News*. https://www.statnews.com/2019/02/05/synthetic-control-arms-clinical-trials/
4. King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis, 9*(2), 137–163. https://doi.org/10.1093/oxfordjournals.pan.a004868
5. Letailleur, V., Chaillol, I., Cherblanc, F., Ghesquières, H., Peyrade, F., Jardin, F., ... & Tessoulin, B. (2023). Synthetic control arm from clinical trials and real-world data from Lysa Group for untreated diffuse large B cell lymphoma patients aged over 80 years: A bona fide strategy for innovative clinical trials. *Blood, 142*(Suppl 1), 72. https://doi.org/10.1182/blood-2023-177708
6. NashBio. (2024, December 18). Synthetic control arms in clinical trials. *Nashville Biosciences Blog*. https://nashbio.com/blog/clinical-trials/synthetic-control-arms-in-clinical-trials/
7. National Bureau of Economic Research. (2016). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *NBER Working Paper No. 22791*. https://www.nber.org/papers/w22791
8. Popat, S., Liu, S. V., Scheuer, N., Hsu, G. G., Lockhart, A., Ramagopalan, S. V., Griesinger, F., & Subbiah, V. (2022). Addressing challenges with real-world synthetic control arms to demonstrate the comparative effectiveness of pralsetinib in non-small cell lung cancer. *Nature Communications, 13*(1), 3500. https://doi.org/10.1038/s41467-022-31193-3
9. Sengupta, S., Ntambwe, I., Tan, K., Liang, Q., Paulucci, D., Castellanos, E., ... & Tiwari, R. (2023). Emulating randomized controlled trials with hybrid control arms in oncology: A case study. *Clinical Pharmacology & Therapeutics, 113*(4), 867–877. https://doi.org/10.1002/cpt.2841
10. Stuart, E. A., et al. (2009). Matching methods for causal inference: A review and a look forward. *Journal of the American Statistical Association, 104*(486), 839–854. https://doi.org/10.1198/jasa.2009.ap08746
11. Gökalp Güneş, S., Gülmez, B., Saçar, K., & Ersoy Şığva, B. İ. (2025). Cardiac tamponade due to internal jugular catheterization after lung resection: A rare but life-threatening complication (case report). *BMC Anesthesiology, 25*, Article 519. https://doi.org/10.1186/s12871-025-03403-5

CONTACT INFORMATION

Author Name: Vasiliki Economou

Company: Pfizer

Email: Vasiliki.Economou@pfizer.com

Author Name: Dimitrios Kounelis

Company: Pfizer

Email: Dimitrios.Kounelis@pfizer.com

Author Name: Nikolaos Kountouris

Company: Pfizer

Email: Nikolaos.Kountouris@pfizer.com