Principal Investigator

First Name: Alejandro
Last Name: Almodóvar Espeso
Degree: MSc Telecommunication Engineering
Primary Affiliation: Universidad Politécnica de Madrid
E-mail: alejandroalmodovar86@gmail.com
State or Province: Madrid
Country: España

General Information

Key Personnel (other than PI):
First Name: Juan
Last name: Parras Moral
Degree: PhD Telecommunication Engineering
Primary Affiliation: Universidad Politécnica de Madrid
SCOPUS ID: 57191201746
Requires Data Access? Yes

First Name: Francisco Javier
Last name: Gómez Fernández-Getino
Degree: BS
Primary Affiliation: Universidad Politécnica de Madrid
SCOPUS ID:
Requires Data Access? Yes

Are external grants or funds being used to support this research?: No external grants or funds are being used to support this research.
How did you learn about the YODA Project?: Scientific Publication

Conflict of Interest

https://yoda.yale.edu/wp-content/uploads/2024/11/YODA_juan.pdf
https://yoda.yale.edu/wp-content/uploads/2024/11/YODA_alejandro.pdf
https://yoda.yale.edu/wp-content/uploads/2024/10/COI-FFG.pdf

Certification

Certification: All information is complete; I (PI) am responsible for the research; data will not be used to support litigious/commercial aims.
Data Use Agreement Training: As the Principal Investigator of this study, I certify that I have completed the YODA Project Data Use Agreement Training

1. NCT01106625 - 28431754DIA3002 - A Randomized, Double-Blind, Placebo-Controlled, 3-Arm, Parallel-Group, Multicenter Study to Evaluate the Efficacy, Safety, and Tolerability of Canagliflozin in the Treatment of Subjects With Type 2 Diabetes Mellitus With Inadequate Glycemic Control on Metformin and Pioglitazone Therapy
2. NCT01137812 - 28431754DIA3015 - A Randomized, Double-Blind, Active-Controlled, Multicenter Study to Evaluate the Efficacy, Safety, and Tolerability of Canagliflozin Versus

Sitagliptin in the Treatment of Subjects With Type 2 Diabetes Mellitus With Inadequate Glycemic Control on Metformin and Sulphonylurea Therapy
3. NCT01106651 - 28431754DIA3010 - A Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Multicenter Study to Evaluate the Efficacy, Safety, and Tolerability of Canagliflozin Compared With Placebo in the Treatment of Older Subjects With Type 2 Diabetes Mellitus Inadequately Controlled on Glucose Lowering Therapy
4. NCT01106677 - 28431754DIA3006 - A Randomized, Double-Blind, Placebo and Active-Controlled, 4-Arm, Parallel Group, Multicenter Study to Evaluate the Efficacy, Safety, and Tolerability of Canagliflozin in the Treatment of Subjects With Type 2 Diabetes Mellitus With Inadequate Glycemic Control on Metformin Monotherapy
5. NCT00968812 - 28431754DIA3009 - A Randomized, Double-Blind, 3-Arm Parallel-Group, 2-Year (104-Week), Multicenter Study to Evaluate the Efficacy, Safety, and Tolerability of JNJ-28431754 Compared With Glimepiride in the Treatment of Subjects With Type 2 Diabetes Mellitus Not Optimally Controlled on Metformin Monotherapy

What type of data are you looking for?: Individual Participant-Level Data, which includes Full CSR and all supporting documentation

Research Proposal

Project Title

Synthetic control arm studies for causal inference evaluation

Narrative Summary:

This study uses machine learning and causal inference to evaluate medical treatments by combining data from randomized controlled trials (RCTs) and observational sources. A "synthetic control arm" is created using observational data to supplement or replace traditional RCT control groups, blending the reliability of RCTs with the cost-efficiency and ethical benefits of observational data. Access to both RCT datasets and observational data is necessary, aligned with key literature. This approach aims to improve clinical trials, reduce costs, and address ethical concerns, ultimately accelerating medical treatment development for public health benefit.

Scientific Abstract:

Background:This study investigates causal inference techniques for synthetic control arms, focusing on methods like propensity weighting and g-computation. These methods offer solutions to the ethical and financial limitations of traditional randomized controlled trials (RCTs). The research builds on foundational studies such as "External Control Arm Analysis: An Evaluation of Propensity Score Approaches, G-Computation, and Doubly Debiased Machine Learning" by Nicolas Loiseau et al.

Objective:The objective is to explore novel causal inference and machine learning techniques not yet applied to synthetic control arms and compare them with established methods to evaluate potential advantages.

Study Design:The study will use an RCT to estimate true average treatment and individual effects, comparing these estimates with those derived from synthetic control arms using causal inference techniques.

Participants:The study will employ datasets used in the research by Nicolas Loiseau et al., leveraging these data sources for evaluating synthetic control arm performance.

Primary and Secondary Outcome Measures:The primary outcome measure will be average and

individual variable effects on HbA1c. We will will assess consistency between synthetic control methods and RCT-based outcomes.

Statistical Analysis: Statistical analysis will compare treatment effect estimates from synthetic control arms with those derived from RCTs, focusing on the robustness of causal inference methodologies. This analysis will be performed in individual and average effects (population level)

Brief Project Background and Statement of Project Significance:

Synthetic control arms offer an effective alternative to traditional RCTs in medical research by using observational data to create virtual control groups. This approach enhances trial efficiency while reducing ethical concerns and costs. Methods such as propensity weighting and g-computation are key to constructing reliable synthetic controls. This research seeks to extend prior work by introducing novel causal inference and machine learning methods into the development of synthetic control arms, aiming to determine whether these innovations provide significant improvements over established methods. If successful, the study could significantly impact clinical trial design by making trials more cost-effective, ethically feasible, and broadly accessible. The methodology involves using several databases (all of them from YODA project), to employ some of them as external control arm to predict support the predictions.

By refining synthetic control methodologies, this research aims to reduce the reliance on expensive and ethically challenging RCTs while maintaining high standards of clinical evidence. The findings will enhance generalizable scientific and medical knowledge, especially concerning the practical application of synthetic controls in clinical settings. The research builds upon key studies, such as the work by Nicolas Loiseau et al., to make medical research more efficient, ethical, and accessible.

References: Loiseau, N., et al. "External Control Arm Analysis: An Evaluation of Propensity Score Approaches, G-Computation, and Doubly Debiased Machine Learning."

Specific Aims of the Project:

The study aims to evaluate the effectiveness of causal inference methods for creating synthetic control arms and compare these outcomes with those from traditional RCTs. Specifically, the project will assess whether novel causal inference and machine learning techniques offer advantages over well-established methods like propensity weighting, g-computation, and doubly debiased machine learning. The evaluation will be conducted using datasets from existing research by Nicolas Loiseau et al.

The primary hypothesis is that novel machine learning-based causal inference methods can improve the accuracy and generalizability of synthetic control arms. If validated, these methods could provide a viable and effective alternative to RCTs in clinical settings, ultimately making medical research more efficient and less reliant on traditional control groups.

Study Design:

Methodological research

What is the purpose of the analysis being proposed? Please select all that apply.

New research question to examine treatment effectiveness on secondary endpoints and/or within subgroup populations

Confirm or validate previously conducted research on treatment effectiveness

Participant-level data meta-analysis

Meta-analysis using only data from the YODA Project

Develop or refine statistical methods

Research on comparison group

Research Methods

Data Source and Inclusion/Exclusion Criteria to be used to define the patient sample for your study:

The inclusion criteria is selected to meet positivity assumption of causal inference, that is, that the subpopulation groups of patients in each study to be similar in the feature space.
Following Loiseau et al., we have the following inclusion criteria for each study:

- NCT01106625: None (all patients selected)
- NCT01137812: None (all patients selected)
- NTC0110665: Age: 55 to 80 y.o
- NCT01106677: None (all patients selected)
- NCT00968812 : 45>=BMI>=2

On the other hand, although our study aims to combine data from multiple studies, all the data that will be used are the requested in this request to YODA project.

Primary and Secondary Outcome Measure(s) and how they will be categorized/defined for your study:

Following Loiseau et al., the primary endpoint is HbA1c in 12 weeks, since that outcome is in all those control trials. Hemoglobin A1c (HbA1c) is a key biomarker used to evaluate long-term blood glucose control in individuals with diabetes. Its measurement reflects the average blood glucose levels over the previous 8--12 weeks due to the lifespan of red blood cells (approximately 120 days).

No secondary outcome to measure treatment effects will be considered.

Main Predictor/Independent Variable and how it will be categorized/defined for your study:

A set of 40 covariates will be used, based in Loiseau et al., named as in their paper. The main predictors are classified as:

Demographic and habits: age, weight, race, tobacco use

Clinical: diastolic blood preasure, systolic blood preasure, LDL, pulse, tryglicerides

Historical data: Previous concomitant medication anti
hyperglycemic, previous concomitant therapy, Concomitant
medication diabetes,

Other Variables of Interest that will be used in your analysis and how they will be categorized/defined for your study:

The other predictors are clinical data from blood analysis and are listed below.

serum albumin: protein for pressure and transport
alkaline phosphatase: enzyme for liver and bone
alanine transaminase: enzyme for liver damage
aspartate transaminase: enzyme for tissue damage
basophils/leukocytes: ratio for immune health
bilirubin: liver health marker

blood urea nitrogen: kidney marker
calcium: mineral for bones and muscles
cholesterol: lipid for cells, heart risk if high
creatine kinase: enzyme for muscle damage
chloride: electrolyte for balance
serum creatinine: kidney function marker
eosinophils: cells for allergies and infections
glomerular filtration rate: kidney filtration estimate
gamma-glutamyl transferase: liver enzyme
blood sugar level: glucose marker
plasma glucose: blood glucose
hemoglobin A1c: 3-month sugar average
HDL cholesterol: "good" cholesterol
hemoglobin: oxygen carrier in blood
potassium: electrolyte for heart
lymphocytes: immune cells
lymphocytes/leukocytes: immune ratio
magnesium: mineral for muscles
neutrophil: infection-fighting cells
phosphate: bone mineral
platelets: clotting cells
protein: blood protein
sodium: hydration electrolyte

Statistical Analysis Plan:

The statistical analysis plan includes descriptive, bivariate, and multivariable analyses, along with advanced causal inference methods. Descriptive statistics will be used to summarize baseline characteristics of participants, while bivariate analyses will identify relationships between treatment and outcome variables. Multivariable models will be employed to adjust for potential confounding factors. Advanced analyses will include propensity score methods (e.g., matching, weighting) to create a balanced comparison group, as well as g-computation to estimate causal effects. Additionally, non-parametric testing will be used to validate results under minimal distributional assumptions, ensuring robustness of treatment effect estimates from both RCT and synthetic control methods.

Software Used:

Python

Project Timeline:

The proposed study is anticipated to start on January 1, 2025, assuming a prompt approval of the data request. The data analysis phase is expected to be completed by April - May, 2025, involving a focused effort to conduct descriptive causal inference analyses. By June, 2025, the manuscript detailing the findings will be drafted and submitted for publication, allowing sufficient time for quality synthesis of results. Finally, the results will be compiled and reported back to the YODA Project by July, 2025, ensuring that the entire process is completed within a six-month period, aligning with the target to finish by July at the latest.

Dissemination Plan:

The target audience includes researchers and professionals involved in clinical trials, healthcare data science, and causal inference, as well as those interested in improving clinical trial design and methodologies. The manuscript will be submitted to journals in the fields of clinical research, medical statistics, or healthcare informatics to reach an audience interested in methodological improvements in clinical studies. Additionally, the findings may be shared at relevant conferences focused on machine learning and healthcare research.

Bibliography:

Loiseau N, Trichelair P, He M, Andreux M, Zaslavskiy M, Wainrib G, Blum MGB. External control arm analysis: an evaluation of propensity score approaches, G-computation, and doubly debiased machine learning. BMC Med Res Methodol. 2022 Dec 28;22(1):335. doi: 10.1186/s12874-022-01799-z. PMID: 36577946; PMCID: PMC9795588.