The YODA Project
Research Proposal Review

The following page contains the final YODA Project review approving this proposal.

The Yale University Open Data Access (YODA) Project
Yale University Center for Outcomes Research and Evaluation (CORE)
195 Church Street, 5th Floor, New Haven, CT 06510

The YODA Project
Research Proposal Review - Final
(Protocol #: 2024-0580)

Reviewers:

- ☒ Nihar Desai
- ☐ Cary Gross
- ☐ Harlan Krumholz
- ☒ Richard Lehman
- ☒ Joseph Ross
- ☐ Joshua Wallach

Review Questions:	**Decision:**
1. Is the scientific purpose of the research proposal clearly described?	Yes
2. Will request create or materially enhance generalizable scientific and/or medical knowledge to inform science and public health?	Yes
3. Can the proposed research be reasonably addressed using the requested data?	Yes, or it's highly likely
4. Recommendation for this data request:	Approve

Comments:

No additional comments

The YODA Project
Research Proposal Review

Revisions were requested during review of this proposal.
The following pages contain the original YODA Project review and the original submitted proposal.

The Yale University Open Data Access (YODA) Project
Yale University Center for Outcomes Research and Evaluation (CORE)
195 Church Street, 5th Floor, New Haven, CT 06510

The YODA Project
Research Proposal Review - Revisions Requested
(Protocol #: 2024-0580)

Reviewers:

- ☒ Nihar Desai
- ☐ Cary Gross
- ☐ Harlan Krumholz
- ☒ Richard Lehman
- ☒ Joseph Ross
- ☐ Joshua Wallach

Review Questions:	**Decision:**
1. Is the scientific purpose of the research proposal clearly described?	Yes
2. Will request create or materially enhance generalizable scientific and/or medical knowledge to inform science and public health?	Yes
3. Can the proposed research be reasonably addressed using the requested data?	Unsure, further clarification from requestor is needed
4. Recommendation for this data request:	Not Approve

Comments:

1. I certainly understand the motivation of the request - to better utilize RWD for evidence generation. The request is a bit vague in terms of the real world data set and what exactly the investigators seek to do. In addition, they are requesting over 40 trials- perhaps a more limited approach to begin (as a proof of concept) would be more reasonable and feasible.

2. The new analytics are not well characterized in the application, please revise.

Principal Investigator

First Name: Brian
Last Name: Gillette
Degree: PhD
Primary Affiliation: Droice Labs
E-mail: brian@droicelabs.com
State or Province: New York
Country: USA

General Information

Key Personnel (other than PI):
First Name: Mayur
Last name: Saxena
Degree: PhD
Primary Affiliation: Droice Labs
SCOPUS ID:
Requires Data Access? No

First Name: Tasha
Last name: Nagamine
Degree: MS
Primary Affiliation: Droice Labs
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?: Colleague

Conflict of Interest

https://yoda.yale.edu/wp-content/uploads/2024/05/YODA_Project_COI_Gillette.pdf
https://yoda.yale.edu/wp-content/uploads/2024/05/YODA_Project_COI_Saxena.pdf
https://yoda.yale.edu/wp-content/uploads/2024/05/YODA_Project_COI_Nagamine.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. NCT02065791 - 28431754DNE3001 - A Randomized, Double-blind, Event-driven, Placebo-controlled, Multicenter Study of the Effects of Canagliflozin on Renal and Cardiovascular Outcomes in Subjects With Type 2 Diabetes Mellitus and Diabetic Nephropathy
2. NCT01032629 - 28431754DIA3008 - A Randomized, Multicenter, Double-Blind, Parallel, Placebo-Controlled Study of the Effects of JNJ-28431754 on Cardiovascular Outcomes in Adult Subjects With Type 2 Diabetes Mellitus

3. NCT02243202 - 28431754OBE2002 - A Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Investigate the Safety and Efficacy of the Co-administration of Canagliflozin 300 mg and Phentermine 15 mg Compared With Placebo for the Treatment of Non-diabetic Overweight and Obese Subjects
4. NCT00236613 - TOPMAT-OBES-001 - A Randomized, Double-Blind, Placebo-Controlled, Multicenter, Parallel Group, Dose-Response Study to Assess the Efficacy and Safety of Topiramate in the Treatment of Patients With Obesity
5. NCT00231608 - TOPMAT-OBMA-001 - The Safety and Efficacy of Topiramate in Male Patients With Abdominal Obesity: A 6-Month Double-Blind, Randomized, Placebo-Controlled Study With a 6-Month Open-Label Extension
6. NCT00231634 - TOPMAT-OBDM-004 - A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel Group Study of the Efficacy and Safety of Topiramate in the Treatment of Obese, Type 2 Diabetic Patients Inadequately Controlled on Sulfonylurea Therapy
7. NCT00231660 - TOPMAT-OBDM-002 - A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study of the Efficacy and Safety of Topiramate in the Treatment of Obese, Type 2 Diabetic Patients Treated With Metformin
8. NCT00236626 - TOPMAT-OBDM-001 - A 9 Month, Double-Blind, Placebo-Controlled Study With a Blinded Crossover Transition to Open-Label Extension, Evaluating the Safety and Effectiveness of Topiramate on Insulin Sensitivity in Overweight or Obese Type 2 Diabetes Patients
9. NCT00231621 - TOPMAT-OBDL-001 - A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel Group, One-year Study of the Efficacy and Safety of Topiramate in the Treatment of Obese Subjects With Dyslipidemia
10. NCT00231647 - TOPMAT-OBD-202 - A Randomized, Double-Blind, Placebo-Controlled, Multicenter, Parallel-Group Study to Assess the Efficacy and Safety of Topiramate OROS Controlled-Release in the Treatment of Obese, Type 2 Diabetic Subjects Managed With Diet or Metformin
11. NCT01989754 - 28431754DIA4003 - A Randomized, Multicenter, Double-Blind, Parallel, Placebo-Controlled Study of the Effects of Canagliflozin on Renal Endpoints in Adult Subjects With Type 2 Diabetes Mellitus
12. NCT00231530 - TOPMAT-OBDM-003 - A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Assess the Safety and Efficacy of Topiramate in the Treatment of Obese, Type 2 Diabetic Patients on a Controlled Diet
13. NCT00231673 - TOPMAT-NP-005 - A Double-Blind, Placebo-Controlled, Parallel Group Study to Evaluate the Effect of Topiramate on Electrophysiological Parameters in Subjects With Diabetic Peripheral Polyneuropathy
14. NCT02025907 - 28431754DIA4004 - A Randomized, Double-blind, Placebo Controlled, 2-arm, Parallel-group, 26-week, 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 Sitagliptin Therapy
15. NCT01340664 - 28431754DIA2003 - 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
16. NCT01381900 - 28431754DIA3014 - A Randomized, Double-Blind, Placebo-Controlled, Parallel Group, 18-Week 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 Alone or in Combination With a Sulphonylurea
17. NCT01809327 - 28431754DIA3011 - A Randomized, Double-Blind, 5-Arm, Parallel-Group, 26-Week, Multicenter Study to Evaluate the Efficacy, Safety, and Tolerability of Canagliflozin in Combination With Metformin as Initial Combination Therapy in the Treatment of Subjects With Type 2 Diabetes Mellitus With Inadequate Glycemic Control With Diet and Exercise
18. 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
19. NCT01106690 - 28431754DIA3012 - 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 Sulphonylurea Therapy

20. 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
21. NCT00210808 - CAPSS-220 - A Multicenter, Randomized, Double-blind, Placebo-controlled, Flexible-dose Study to Assess the Safety and Efficacy of Topiramate in the Treatment of Moderate to Severe Binge-eating Disorder Associated With Obesity
22. NCT00650806 - 28431754OBE2001 - A Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Dose-Ranging Study to Investigate the Safety and Efficacy of JNJ-28431754 in Nondiabetic Overweight and Obese Subjects
23. 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
24. 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
25. NCT01081834 - 28431754DIA3005 - A Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Multicenter Study to Evaluate the Efficacy, Safety, and Tolerability of Canagliflozin as Monotherapy in the Treatment of Subjects With Type 2 Diabetes Mellitus Inadequately Controlled With Diet and Exercise
26. NCT01064414 - 28431754DIA3004 - A Randomized, Double-Blind, Placebo-Controlled, 3-Arm, Parallel-Group, 26-Week, Multicenter Study With a 26-Week Extension, to Evaluate the Efficacy, Safety and Tolerability of Canagliflozin in the Treatment of Subjects With Type 2 Diabetes Mellitus Who Have Moderate Renal Impairment
27. 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
28. NCT03267576 - 28431754DIA4026 - Canagliflozin Continuous Glucose Monitoring (CANA CGM) Trial: A Pilot Randomized, Double-Blind, Controlled, Crossover Study on the Effects of the SGLT-2 Inhibitor Canagliflozin (vs. the DPP-4 Inhibitor Sitagliptin) on Glucose Variability in Mexican Patients With Type 2 Diabetes Mellitus Inadequately Controlled on Metformin
29. NCT02139943 - 28431754DIA2004 - A Randomized Phase 2, Double-blind, Placebo-controlled, Treat-to-Target, Parallel-group, 3-arm, Multicenter Study to Assess the Efficacy and Safety of Canagliflozin as Add-on Therapy to Insulin in the Treatment of Subjects With Type 1 Diabetes Mellitus
30. NCT01385202 - Smart-AF - THERMOCOOL® SMARTTOUCH™ Catheter for the Treatment of Symptomatic Paroxysmal Atrial Fibrillation
31. NCT02382016 - AC-055-404 - A Randomized, Double-blind, Placebo-controlled, Prospective, Multicenter, Parallel Group Study to Assess the Safety and Efficacy of Macitentan in Patients With Portopulmonary Hypertension
32. NCT03078907 - AC-065A404 - A Multi-center, Double-blind, Placebo-controlled Phase 4 Study in Patients With Pulmonary Arterial Hypertension to Assess the Effect of Selexipag on Daily Life Physical Activity and Patient's Self-reported Symptoms and Their Impacts
33. NCT02471183 - AC-065A304 - Multicenter, Open-label, Single-group Study to Assess the Tolerability and the Safety of the Transition From Inhaled Treprostinil to Oral Selexipag in Adult Patients With Pulmonary Arterial Hypertension
34. NCT02070991 - AC-055G201 - A Prospective, Multicenter, Double-blind, Randomized, Placebo-controlled, Parallel-group, 12-week Study to Evaluate the Safety and Tolerability of Macitentan in Subjects With Combined Pre- and Post-capillary Pulmonary Hypertension (CpcPH) Due to Left Ventricular Dysfunction
35. NCT00313222 - AC-052-366 - Prospective, Randomized, Placebo-controlled, Double-blind, Multicenter, Parallel Group Study to Assess the Efficacy, Safety and Tolerability of Bosentan in Patients With Inoperable Chronic Thromboembolic Pulmonary Hypertension (CTEPH)

36. NCT00660179 - AC-055-302 - A Multicenter, Double-blind, Randomized, Placebo-controlled, Parallel Group, Event-driven, Phase III Study to Assess the Effects of Macitentan (ACT-064992) on Morbidity and Mortality in Patients With Symptomatic Pulmonary Arterial Hypertension
37. NCT00091715 - AC-052-364 - A Randomized, Double-blind, Placebo-controlled, Multicenter Study to Assess the Efficacy, Safety, and Tolerability of Bosentan in Patients With Mildly Symptomatic Pulmonary Arterial Hypertension (PAH)
38. NCT00303459 - AC-052-414 (COMPASS-2) - Effects of Combination of Bosentan and Sildenafil Versus Sildenafil Monotherapy on Morbidity and Mortality in Symptomatic Patients With Pulmonary Arterial Hypertension - A Multicenter, Double-blind, Randomized, Placebo-controlled, Parallel Group, Prospective, Event Driven Phase IV Study
39. NCT00319111 - AC-052-370 (BENEFIT OL) - Long-term Open-label Extension Study in Patients With Inoperable Chronic Thromboembolic Pulmonary Hypertension (CTEPH) Who Completed Protocol AC-052-366 (BENEFIT)
40. NCT01106014 - AC-065A302 - A Multicenter, Double-blind, Placebo-controlled Phase 3 Study Assessing the Safety and Efficacy of Selexipag on Morbidity and Mortality in Patients With Pulmonary Arterial Hypertension
41. NCT00236665 - TOPMAT-OBHT-001 - A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel Group Study of the Efficacy and Safety of Topiramate in the Treatment of Obese Patients With Mild to Moderate Essential Hypertension
42. NCT00642278 - 28431754DIA2001 - A Randomized, Double-Blind, Placebo-Controlled, Double-Dummy, Parallel Group, Multicenter, Dose-Ranging Study in Subjects With Type 2 Diabetes Mellitus to Evaluate the Efficacy, Safety, and Tolerability of Orally Administered SGLT2 Inhibitor JNJ-28431754 With Sitagliptin as a Reference Arm
43. NCT00816166 - VISSIT CA-2007-01 - Phase III Study of Pharos Vitesse Neurovascular Stent System Compared to Best Medical Therapy for the Treatment of Ischemic Disease
44. NCT00236639 - TOPMAT-OBES-002 - A Randomized, Double-Blind, Placebo-Controlled, Multicenter, Parallel Group, Dose-Response Study to Assess the Efficacy and Safety of Topiramate in the Treatment of Patients With Obesity
45. NCT00236600 - TOPMAT-OBES-004 - A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study of Efficacy and Safety of Topiramate in Weight Loss Maintenance in Obese Patients Following Participation in an Intensive, Non-Pharmacologic Weight Loss Program

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

Research Proposal

Project Title

Developing approaches to accelerate clinical trials using real world data (RWD) from electronic health records (EHR)

Narrative Summary:

The use of real-world data (RWD) from electronic health records (EHR) in clinical trials stands to significantly improve the speed and efficiency of developing new therapies. However, data quality/reliability issues in EHR-derived study data can lead to biases that confound the exposure-outcome relationship and limit comparability with clinical trial data. This project will develop methods to leverage RWD in randomized controlled trials (RCT) while accounting for RWD data quality/reliability issues by using a novel data traceability solution to measure error in EHR-derived data quantitatively evaluate impact of biases in comparative outcome analyses.

Scientific Abstract:

Background: Randomized controlled trials (RCTs) are the gold standard for evidence generation but are slow, expensive, and in certain circumstances ethically challenging. Real-world data (RWD) from electronic health records (EHR) could improve efficiency of RCTs for multiple use cases, including external control arms (ECAs) when randomization is not ethical/feasible, hybrid controls that collect RWD for patients in the trial, and/or RWD-based prognostic models that increase statistical power through covariate adjustment. However, FDA is rejecting RWD for primary evidence due to lack of demonstrated reliability - accuracy, completeness, and traceability – as outlined in FDA guidance.

Objective: This project will develop methods to utilize RWD in RCTs with quantified reliability (accuracy, completeness, and traceability) that is used to account for biases resulting from RWD data loss and transformation errors in comparative outcome analyses.

Study Design: Methodological research on integrated analyses of RCT and RWD.

Participants: RCTs will be selected from the YODA project and matched cohorts from RWD will be selected for each RCT from Droice Labs' US hospital partners.

Primary and Secondary Outcome Measure(s): The same outcome measures from each selected RCT will be used in the RCT/RWD analysis.

Statistical Analysis: For each selected RCT, analysis will assess the replicability of the RCT inferences with RWD comparator arms and quantify the stability of inferences under measured misclassification in RWD using probabilistic quantitative bias analysis.

Brief Project Background and Statement of Project Significance:

RCTs are the gold standard for evidence generation for new treatments but are highly resource intensive. If RWD from EHR could be used for evidence generation at scale, it could dramatically accelerate therapy development and lead to improved care in wide-ranging therapeutic areas. For example, RWD external control arms (ECAs) can be used when randomization is not ethical/feasible1–8. For common conditions for which FDA requires RCTs, hybrid controls can capture major portions of data from RWD to reduce the cost and time burden of manual data collection9–11, and prognostic disease progression models developed from RWD can be applied to increase statistical power through covariate adjustment12.

However, there are major challenges in using EHR data for generating reliable inferences in clinical trials, because EHR data is highly messy and noisy compared to RCT data and typically requires complex data transformation steps to prepare it for analysis1,2,4,13–16. Transformation errors and data loss in RWD processing induce misclassifications and subsequent biases in the analysis data that can render it unreliable for critical regulatory decisions because the potential impact of such biases on study inferences cannot be quantitatively assessed17. Accordingly, FDA has been rejecting submissions using RWD because data reliability - accuracy, completeness, and traceability – has not been sufficiently demonstrated according to FDA guidance6,13,17,18. For RWD to be reliable, traceability to raw data across all data transformations is required to validate source data and quantify the accuracy of data transformations.

Droice Labs developed SuperLineage, an element-level lineage solution specifically designed to solve the challenges of traceability and validation of RWD transformations. SuperLineage losslessly captures source EHR in a standardized format, which is critical for measuring information loss and error. SuperLineage is used both to find and fix data loss and transformation errors to improve data quality and to measure transformation performance (e.g. sensitivity and sensitivity of computational phenotyping algorithms17) so that error bounds can be quantitatively accounted for in comparative efficacy/safety analyses. Droice Labs' discussed SuperLineage with the FDA, where both FDA and Droice agreed that accuracy, completeness, and traceability are necessary for regulatory use of RWD19. Furthermore, Droice and CDISC, the body responsible for developing regulatory data submission standards for FDA, are building an RWD lineage metadata standard based on

SuperLineage for FDA regulatory submissions20,21.

In this project, patient-level data from a variety of RCTs will be integrated with matched cohorts from RWD from multiple US hospital partners of Droice Labs. SuperLineage will be used to quantify accuracy and completeness of RWD variables while maintaining required regulatory-grade traceability. For each selected RCT, the primary analysis will be replicated across the RCT and RWD study arms to evaluate whether the base inferences of the RCT are reproducible in the RWD with and without considering the error bound quantified through SuperLineage.

Specific Aims of the Project:

Aim 1: Develop high-quality EHR-derived RWD cohorts matched to RCT treatment and control arms for a variety of cardiometabolic indications. Data reliability (accuracy, completeness, and traceability) of all EHR-derived study variables (inclusion/exclusion criteria, exposures, covariates, and outcomes) will be quantitatively evaluated using Droice SuperLineage.

Aim 2: Evaluate reproducibility of efficacy and safety endpoints across RCT and RWD arms for each selected RCT with and without considering the data reliability from Aim 1.

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

Participant-level data meta-analysis

Meta-analysis using data from the YODA Project and other data sources

Develop or refine statistical methods

Research on clinical trial methods

Research Methods

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

Clinical trial data: Phase 2-4 RCT from the YODA project are selected for the following cardiometabolic conditions/indications: heart failure, myocardial infarction, heart valve disease, atrial fibrillation, stroke, chronic kidney disease, diabetes (type 1 and 2), obesity, hypertension (essential and pulmonary), NASH, and hyperlipidemia. All patients from each trial will be included (no exclusion criteria).

EHR data: Deidentified EHR data from relevant hospitals in the Droice data warehouse will be used to select patients that match each RCT based on the inclusion/exclusion specific to that trial. We will upload deidentified patient-level analysis data (i.e. CSVs with treatment status, baseline covariate values, and outcome status/values as columns) derived from EHR data to the YODA platform matched to each RCT. Research use of the deidentified EHR data has been verified IRB exempt as non-human subjects research not requiring informed consent according to 45CFR46.104(d) (Solutions IRB).

Pooled IPD analysis from RCT and EHR data will be performed on the YODA project platform.

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

For each selected RCT, the analysis will compare the same primary/secondary outcomes as the RCT for EHR-derived study cohorts, subject to the availability of the corresponding outcome data in the RWD source for a given EHR-derived comparator cohort.

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

For each selected RCT, the analysis will compare the same intervention/standard of care as the RCT for the EHR-derived study cohorts, subject to the availability of the corresponding intervention/standard of care in the RWD source for a given EHR-derived comparator cohort.

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

For each selected RCT, the analysis will utilize the same covariates as the RCT for the EHR-derived study cohorts, subject to the availability of the corresponding covariates in the RWD source for a given EHR-derived comparator cohort.

Statistical Analysis Plan:

For each selected RCT, analysis will include arms from the RCT and matched arms from RWD:
1) The treatment arm(s) from the RCT
2) The control arm from the RCT
3) Matched treatment arm(s) from RWD with patients exposed to the RCT intervention(s)
4) Matched control arm from RWD treated with the RCT SoC

Outcomes for treatments and standard of care will be compared across all arms:
1) RCT treatment(s) vs. RWD SoC to investigate clinical trial treatment efficacy vs. real-world SoC (representing an external control arm study)
2) RCT treatment(s) vs. RWD treatment(s) to investigate differential efficacy in RCT vs. real-world treatment
3) RWD treatment(s) vs. RWD SoC to investigate real-world treatment effectiveness
4) RCT SoC vs. RWD SoC to investigate differential efficacy of SoC in RCT vs. real-world SoC

Creation of RCT matched cohorts in RWD:
Inclusion/exclusion criteria for each RCT will be applied to RWD to generate an eligible patient pool for matching. Propensity score and Mahalanobis distance matching methods will be used to create matched cohorts between RCT and RWD cohorts22–24.

Quantitative bias analysis (QBA) to evaluate inference robustness to EHR variable misclassification errors:
First, base inferences will be calculated to determine if RCT results are reproduced in RWD without considering misclassification in RWD. Then, with misclassification assumed negligible in the clinical trial data (due to rigorous data collection and quality assurance protocols), effects of EHR-derived variable misclassification will be systematically investigated through probabilistic quantitative bias analysis (QBA)16,25–28 to determine the stability of the base inferences considering the measured misclassification (sensitivity and specificity) in EHR variables.

QBA will account for misclassification across all study variables, including:
Inclusion/exclusion criteria: misclassified inclusion/exclusion criteria can alter which patients are included in the study cohort, potentially imparting selection bias
Exposures: misclassifications in exposures change which arm (exposed or unexposed) an individual is assigned to, biasing the exposure-outcome relationship
Covariates: misclassifications in prognostic covariates distort the real profiles of confounding factors,

which can bias the control for confounding, making covariate adjustment ineffective or inaccurate Outcomes: misclassifications in outcomes can change the efficacy and safety profile across study groups, biasing the exposure-outcome relationship

Software Used:

Python

Project Timeline:

Estimated start date: July 2024
Estimated analysis completion: July 2025
Estimated manuscript submission: October 2025

Dissemination Plan:

This project is expected to result in conference presentations and at least one major publication in a journal targeting a broad audience interested in advancements in the use of RWD for enhancing clinical evidence generation, such as Nature Scientific Reports, JAMA Network Open, BMC Medical Research Methodology, etc.

Bibliography:

1. Kim TE, Park SI, Shin KH. Incorporation of real-world data to a clinical trial: use of external controls. *Transl Clin Pharmacol*. 2022;30(3):121. doi:10.12793/TCP.2022.30.E14
2. Yap TA, Jacobs I, Baumfeld Andre E, Lee LJ, Beaupre D, Azoulay L. Application of Real-World Data to External Control Groups in Oncology Clinical Trial Drug Development. *Front Oncol*. 2021;11:695936. doi:10.3389/fonc.2021.695936
3. Zou KH, Vigna C, Talwai A, et al. The Next Horizon of Drug Development: External Control Arms and Innovative Tools to Enrich Clinical Trial Data. *Ther Innov Regul Sci*. 2024;58(3):443-455. doi:10.1007/s43441-024-00627-4
4. Carrigan G, Bradbury BD, Brookhart · M Alan, et al. External Comparator Groups Derived from Real-world Data Used in Support of Regulatory Decision Making: Use Cases and Challenges. *Current Epidemiology Reports 2022 9:4*. 2022;9(4):326-337. doi:10.1007/S40471-022-00305-9
5. Yin X, Mishra-Kalyan PS, Sridhara R, Stewart MD, Stuart EA, Davi RC. Exploring the Potential of External Control Arms created from Patient Level Data: A case study in non-small cell lung cancer. *J Biopharm Stat*. 2022;32(1):204-218. doi:10.1080/10543406.2021.2011901
6. FDA. Designing Sound Clinical Trials That Incorporate Real-World Data | FDA. Accessed May 29, 2024. https://www.fda.gov/drugs/regulatory-science-action/designing-sound-clinical-trials-incorporate-real-world-data
7. Jahanshahi M, Gregg K, Davis G, et al. The use of external controls in FDA regulatory decision making. *Ther Innov Regul Sci*. 2021;55(5):1019-1035. doi:10.1007/s43441-021-00302-y
8. Jiao F, Chen YF, Min M, Jimenez S. Challenges and potential strategies utilizing external data for efficacy evaluation in small-sized clinical trials. *J Biopharm Stat*. 2022;32(1):21-33. doi:10.1080/10543406.2021.2011906
9. Hampson L V., Izem R. Innovative Hybrid Designs and Analytical Approaches Leveraging Real-World Data and Clinical Trial Data. *Real-World Evidence in Medical Product Development*. Published online January 1, 2023:211-232. doi:10.1007/978-3-031-26328-6_12
10. Zhu M, Sridhar S, Hollingsworth R, et al. Hybrid clinical trials to generate real-world evidence: design considerations from a sponsor's perspective. *Contemp Clin Trials*. 2020;94. doi:10.1016/J.CCT.2019.105856
11. Ventz S, Khozin S, Louv B, et al. The design and evaluation of hybrid controlled trials that leverage external data and randomization. *Nature Communications 2022 13:1*. 2022;13(1):1-11. doi:10.1038/s41467-022-33192-1
12. Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products | FDA. Accessed June 9, 2024. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adjusting-covariates-randomized-clinical-trials-drugs-and-biological-products

13. FDA. Considerations for the Use of Real-World Data and Real-World Evidence To Support Regulatory Decision-Making for Drug and Biological Products | FDA. Accessed May 29, 2024. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug
14. Liu F, Demosthenes P. Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Medical Research Methodology 2022 22:1*. 2022;22(1):1-10. doi:10.1186/S12874-022-01768-6
15. Gray CM, Grimson F, Layton D, Pocock S, Kim J. A Framework for Methodological Choice and Evidence Assessment for Studies Using External Comparators from Real-World Data. *Drug Saf*. 2020;43(7):623-633. doi:10.1007/S40264-020-00944-1
16. Gray C, Ralphs E, Fox MP, et al. Use of quantitative bias analysis to evaluate single-arm trials with real-world data external controls. *Pharmacoepidemiol Drug Saf*. 2024;33(5). doi:10.1002/pds.5796
17. FDA. Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products | FDA. Accessed May 30, 2024. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory
18. FDA. Data Standards for Drug and Biological Product Submissions Containing Real-World Data | FDA. Accessed May 30, 2024. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-standards-drug-and-biological-product-submissions-containing-real-world-data
19. Droice Labs Discusses SuperLineage for Real-World Data Traceability at FDA Listening Session. Accessed May 29, 2024. https://finance.yahoo.com/news/droice-labs-discusses-superlineage-real-150000597.html?guccounter=1
20. Nagamine T, Hume S. Transforming RWD for Regulatory Submissions: How to Use SDTM for RWD. In: *PHUSE US Connect*. ; 2023.
21. RWD Lineage | CDISC. Accessed May 30, 2024. https://www.cdisc.org/rwd-lineage
22. King G, Nielsen R. Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*. 2019;27(4):435-454. doi:10.1017/PAN.2019.11
23. Ripollone JE, Huybrechts KF, Rothman KJ, Ferguson RE, Franklin JM. Implications of the Propensity Score Matching Paradox in Pharmacoepidemiology. *Am J Epidemiol*. 2018;187(9):1951-1961. doi:10.1093/AJE/KWY078
24. Ali MS, Prieto-Alhambra D, Lopes LC, et al. Propensity score methods in health technology assessment: principles, extended applications, and recent advances. *Front Pharmacol*. 2019;10:973. doi:10.3389/fphar.2019.00973
25. Fox MP, MacLehose RF, Lash TL. Applying Quantitative Bias Analysis to Epidemiologic Data. Published online 2021. doi:10.1007/978-3-030-82673-4
26. Petersen JM, Ranker LR, Barnard-Mayers R, Maclehose RF, Fox MP. A systematic review of quantitative bias analysis applied to epidemiological research. *Int J Epidemiol*. 2021;50(5):1708-1730. doi:10.1093/ije/dyab061
27. Fox MP, Maclehose RF, Lash TL. SAS and R code for probabilistic quantitative bias analysis for misclassified binary variables and binary unmeasured confounders. *Int J Epidemiol*. 2023;52(5):1624-1633. doi:10.1093/ije/dyad053
28. Thorlund K, Duffield S, Popat S, et al. Quantitative bias analysis for external control arms using real-world data in clinical trials: a primer for clinical researchers. *J Comp Eff Res*. 2024;13(3). doi:10.57264/cer-2023-0147