OPERAND Research Application

Project Number	***To be assigned by OptumLabs upon submission***
Project Title	Emulating target trials using observational studies: use case in OptumLabs
Principal Investigator	Issa Dahabreh, MD MS
Institution	Brown University & Harvard University
Email	issa_dahabreh@brown.edu

All participants on the project team are subject to OptumLabs® policies, procedures, and all agreements between your institution and OptumLabs. Before sharing any feasibility results with non-institution collaborators, please ensure they are covered by confidentiality terms with OptumLabs (NDA or corresponding confidentiality terms must be in place). This document will be used for project oversight, accountability, coordination of activities and decision making. In the event of a conflict between this document and any agreements between your institution and OptumLabs, the terms and conditions of any such agreements shall control.

Research Application Completion Instructions

For Research Application Review:

- Review and update the cover page
- Review sections 1-2
- Complete Sections 3-15
- Please fully read Sections 16-18

The Research Application is intended to provide study protocol for assessment of feasibility and alignment with OptumLabs' mission by OptumLabs Research team staff members. It is understood that the study protocol is a living document and research designs (such as planned study variables or statistical analyses) may change as projects progress. OptumLabs recommends that Research Applications be annotated with changes throughout the life of the project so that an accurate account of data set construction, variable operationalization, and statistical analysis is maintained. Research Applications are reviewed by the OptumLabs Research Review Committee (RRC). More information on this review can be found on The Bridge.

We recognize that certain Partners may have existing formats for detailed study protocols. Partners may submit material requested in Section 3 in their own format as a supplement to the application, provided that: 1) all of the content requested in the application is included; and 2) all other application sections are completed and submitted along with the supplement.

1. Project Summary

a. Problem Formulation / Hypothesis

While prospectively planned clinical trial data remains the cornerstone of regulatory submissions, to date real world evidence (RWE) has not been routinely utilized or accepted for drug approval due to a number of challenges. Recent interest in RWE studies has been driven by the desire to bring innovative products to patients more quickly than the traditional drug development path involving RCTs. This desire has been expressed through regulatory mandates. The Observational Patient Evidence for Regulatory Approval and uNderstanding Disease (OPERAND) program is designed to better inform the use of RWE from retrospective observational studies in medicine and regulatory decision-making. We propose a program that will first use retrospective observational data to determine if such data can confirm previously published RCT results and, if successful, extend the use of the data for other potential applications.

b. Disease Area and Project Type Categorization

OptumLabs tracks information about projects conducted by our Partners for the purpose of examining trends in research. Please help us with these efforts by answering the two questions below:

Please select one "project type" or theme that best describes your project. We recognize that categories are not mutually exclusive, and your project may fall into several categories.

Methods and Innovation in Observational Data Analysis

Please select the disease area which best represents the focus of your project. If multiple disease areas are included in your project, please select the disease area that is most prevalent or you feel is of most interest. Chapters of the ICD10 coding system are provided to aid your selection.

Not Applicable

c. Study Purpose and Objectives

The purpose of this study to replicate two previously published randomized controlled trials of pharmacological products that were used as the basis of marketing approval by the FDA, the

ROCKET Atrial Fibrillation study and the LEAD-2 study. For each trial, the initial objective is to mimic the inclusion/exclusion criteria, endpoint definitions, exposure windows, and other design features of each study as closely as possible. Then, using a series of multivariate methods, average treatment effect (ATE) estimates will be produced and compared to those reported in the original publication.

The specific objectives for this project are:

1. Mimic the inclusion/exclusion criteria of the trial, as well as the primary endpoint, and follow-up period. Carefully document your definitions as one of the goals of OPERAND is to understand how differences in researcher decision-making may introduce variability into ATE estimates.
2. Create three patient cohorts for each study: (1) using the full sample of patients with claims data, irrespective of whether they have linked EHR data, (2) the sample of claims that link with EHR data, and (3) the sample of patients with claims plus EHR data. (See discussion below on comparisons of these samples.)
3. Perform multivariate analyses to estimate ATEs using the following:
    - A single equation method such as generalized linear models (GLM) to introduce the treatment variable as a covariate. Please describe your criteria/method for selecting the other covariates included in the model.
    - Propensity score matching. We leave the specification of variables to be included in the propensity score model, as well as other decisions, such as the size of the calipers, to the research team but please describe these decisions and criteria fully.
    - Inverse probability weights. Same documentation requirements as the other two methods.
    - At least one other appropriate method at the discretion of the researcher (e.g., regression discontinuity methods, instrumental variables, differences in differences, G-estimation, targeted maximum likelihood estimation (TLME)). More than one method is encouraged.
4. Estimate separate sets of models for the three samples of patients. Linking to EHR will substantially reduce the sample but provide additional covariate controls. Use bootstrapping methods for all ATE estimates from the observational data to evaluate the effects of alternative methods and data samples on estimation bias (relative to the published RCT ATE) and standard errors.
5. How do the ATE estimates from the 3 samples compare?
6. Compare the results from the multivariate analyses to the published estimates of ATE using two methods:
    - Regulatory agreement—defined as statistically significant result with directional equivalence between the RCT and observational study.
    - Estimate agreement—defined as the point estimate of the observational study falling within the 95% confidence interval of the ATE from the RCT

using the reported standard errors of the RCT to define the confidence interval.

7. Are the results similar? If not, why do you think they differ? Do you have greater confidence in some of the observational results than others? Why or why not?

d. Data Sources

We will use the following currently available data from 1994-2018. Note that for the final primary analyses we will probably restrict the range to years during which all compared interventions (for each trial emulation) were approved for the indication evaluated in the emulated trial. That is to say, for the LEAD-2 trial emulation, we will use data from 2010 onwards and for the ROCKET trial emulation we will use data from 2011 onwards. For the final emulation analyses and the "extended population" analyses (if requested) we will update the year range as 2019 data become available in OptumLabs.

☒ Administrative claims data: Enrollment files
☒ Administrative claims data: Medical claims transactions
☒ Administrative claims data: Pharmacy claims transactions
☒ PanTher (EHR-derived)
☒ Laboratory results – claims-sourced/health plan affiliated
☒ Supplemental Oncology Data
☒ Race/ethnicity
☒ Mortality Status
☒ Cost data: insurer-paid amounts
☒ Cost data: patient-paid amounts
☒ Benefits Design
☐ Patient distance traveled
☒ Socioeconomic Status Data
☒ Health Risk Assessment (HRA)
☒ Consumer/Lifestyle data
☒ RUCA codes
☐ Area Health Resource Files (AHRF)
☐ American Hospital Association (AHA) Data
☒ Expanded discharge status field

Based on the above and the information in the Research Application Appendix, this study will utilize the following view (select one):

☒ **NATIONAL**
☐ **UNIFIED**
☐ **LIFESTYLE**
☐ **SES**

☐ **STATE**
☐ **MORTALITY**
☐ **COUNTY**
☐ **ZIP5**

Study subjects will include:

☒ Commercial enrollees
☒ Medicare Advantage enrollees

2. Research Plan

a. Population, study cohorts, and special subgroups of interest

Trial emulations: **Table 1** summarizes our plan for the emulation of the ROCKET trial and **Table 2** summarizes our plan for the LEAD-2 trial. We now provide some additional information on cross-cutting issues that apply to both emulations.

Cohort construction: We will follow the recommendation in the RFP regarding the strategy for cohort construction. Specifically, we will start with the claims data first and then match the claims cohort to the subset of patients who also have EHR data. We will compare the results of three cohort creation and data selection strategies: (1) patients with claims irrespective of EHR status; (2) patients with claims who also have EHR data available but using only data from the claims subset; (3) patients with both claims and EHR data. We do note that we followed this strategy for our PCORI-funded Methods Research Award that used OptumLabs data to perform a number of trial emulations (see **Section 7** for additional information).

Eligibility criteria: To determine eligibility for our emulations, we will rely heavily on diagnostic and procedure codes from claims or EHR data to operationalize the trial inclusion and exclusion criteria. Whenever possible, we will implement validated algorithms that have been used in prior studies of similar data (we discuss this further in **Section 3b**). We will attempt to emulate the study design of the emulated trials to the extent possible. For example, in the ROCKET trial emulation we will implement a new-user (incident user) designs[1-3] and in the LEAD-2 trial we will emulate the run-in design using the methods described previously by Hernan and Robins.[4]

Continuous enrollement (pre-baseline): To implement the trial eligibility criteria and be able to extract baseline covariate values for statistical analyses, we will require that patients remained continuously enrolled in their plan for at least 6 months before baseline. Note that only a subset of a patient's medical encounters are captured in EHR data, even if the patient remains continuously enrolled in the same insurance plan, because patients may choose to receive care by providers who do not contribute data to OptumLabs.[5]

Covariate assessment period (pre-baseline): In all emulations, we will collect covariate information over the 6 months before the start of treatment. When multiple laboratory measurements are available in that period (e.g., in emulations using both claims and EHR data), we will use the values most proximal to baseline.

Enrollment and followup periods: We anticipate limiting the period during which individuals in the data can be eligible for our emulations to the periods when all the compared interventions were available. We will emulate the baseline time (time zero) choice and followup duration of the

emulated trials (see **Tables 1** and **2** for details). We will not require a minimum follow-up duration to avoid introducing selection bias.[6] We will follow individuals until the end of followup in the corresponding emulated trial, the occurrence of an outcome event, death, or loss to follow-up, whichever occurred first. Loss to follow-up will be defined as plan dis-enrollment from the claims data.

Expanded population analyses (pending TEP approval): If the conduct of the expanded population analyses is requested by the TEP, we will use the same specifications as for the emulations described above (and in **Tables 1** and **2**), except for the eligibility criteria, which will be systematically relaxed. Our understanding, based on the RFP, is that the TEP will provide input for how to systematically "expand the aperture". For qualitative restrictions, the criteria can be broadened by simply removing one or more restrictions (e.g., by allowing individuals at low risk of stroke in the rivaroxaban analyses, even though such individuals will have been excluded from the ROCKET trial emulation). For quantitative selection criteria (e.g., age), the criteria can be broadened gradually (e.g., require age <80 years, <85 years, <90 years in the LEAD-2 emulation). Natural as these choices for relaxing the eligibility criteria may appear, they create a multivariate "criteria relaxation surface" that cannot be exhaustively explored (once the number of criteria being relaxed grows above a small number, the size of the space of possible observational studies becomes very large). We will reduce the dimensionality of the space by proposing "bundles" of criteria that define 3-5 new target patient populations, that is to say, we will combine versions of the criteria in such a way as to induce a spectrum ranging from a "highly selected" population (often, but not necessarily, this will coincide with the actual trial emulation) to a "minimally selected/unselected" population. These bundles of criteria will be provided to the TEP for review along with the other deliverables on September 2019.

We intend to use the same analysis plan for the expanded population analyses as for the emulation analyses. Please see **Section 11** for details of our analysis plan.

Table 1: Main aspects of the ROCKET trial emulation.[7]

Component of emulation	ROCKET trial emulation (criteria abstracted from the trial protocol)	Notes for the operationalization of the eligibility criteria in OptumLabs (see also our remarks regarding variable definitions in Section 3b)
Eligibility criteria	– Age ≥18 years – Non-valvular atrial fibrillation, as documented on electrocardiography. Atrial fibrillation must be documented by ECG evidence (e.g., 12-lead ECG, rhythm strip, Holter, pacemaker interrogation) within 30 days before randomization. In addition, subjects must have medical evidence of atrial fibrillation within 1 year before and at least one day before the qualifying ECG evidence. This could be obtained from the subject's medical record. – Moderate-to-high risk for stroke. Elevated risk was indicated by a history of stroke, transient ischemic attack, or systemic embolism or at least two of the following risk factors: heart failure or a left ventricular ejection fraction of 35% or less, hypertension, age ≥75 years, or the presence of diabetes mellitus (i.e., a CHADS2 score of 2 or more, on a scale ranging from 1 to 6, with higher scores indicating a greater risk of stroke). – The proportion of patients who had not had a previous ischemic stroke, transient ischemic attack, or systemic embolism and who had no more than two risk factors was limited to 10% of the cohort for each region; the remainder of patients were required to have had either previous thromboembolism or three or more risk factors. – *Exclusion Criteria:* hemodynamically significant mitral valve stenosis; prosthetic heart valve (annuloplasty with or without prosthetic ring, commissurotomy and/or valvuloplasty are permitted); planned cardioversion (electrical or pharmacological); transient atrial fibrillation caused by a reversible disorder (e.g., thyrotoxicosis, PE, recent surgery, MI); known presence of atrial myxoma or left ventricular thrombus; active endocarditis; active internal bleeding; history of or condition associated with increased bleeding risk including; planned invasive procedure with potential for uncontrolled bleeding, including major surgery; platelet count <90,000/μL at the screening visit; sustained uncontrolled hypertension: systolic blood pressure ≥180 mmHg or diastolic blood pressure ≥100 mmHg; severe, disabling stroke within 3 months or any stroke within 14 days before the randomization visit; transient ischemic attack within 3 days before the randomization visit; indication for anticoagulant therapy for a condition other than atrial fibrillation (e.g., VTE); treatment with aspirin, intravenous antiplatelets within 5 days before randomization, fibrinolytics within 10 days before randomization; anti-inflammatory drug; systemic treatment with a strong inhibitor of cytochrome P450 3A4 within 4 days before randomization, or planned treatment during the time period of the study; treatment with a strong inducer of cytochrome P450 3A4, within 4 days before randomization, or planned treatment during the time period of the study; anemia; anemia (hemoglobin <10 g/dL) at the screening visit; pregnancy or breast-feeding; any other contraindication to warfarin; known HIV infection at time of screening; calculated CLCR <30 mL/min at the screening visit; known significant liver disease; serious concomitant illness associated with a life expectancy of less than 2 years.	– Age criteria will be implemented using the corresponding claims or EHR field. – Non-valvular atrial fibrillation will be defined using inpatient or outpatient claims, not electrocardiography. All other valvular conditions described in the ROCKET trial's protocol (available as Appendix to the NEJM article reporting the trial results) will be reflected in the routinely collected data using established code algorithms. – As per the RFP, we will not use left ventricular ejection fraction as an eligibility criterion in our main analysis. We may opt to use it in a stability analysis (see paragraph on "Stability analyses"), but we will base our judgement on the completeness of the available data in the EHR. – We will obtain a claims-based version of the CHADS2 score using methods analogous to those described previously.[8] – Conditions such as hypertension, diabetes, recent stroke or transient ischemic attack, indications for anticoagulation other than atrial fibrillation (e.g., recent venous thromboembolism), anemia, significant liver disease, etc., will be coded using established diagnostic algorithms (e.g., as for our prior work using OptumLabs data), combining diagnosis codes and medication history information. – The criterion about centers having up to 10% of patients who had not had a previous ischemic stroke, transient ischemic attack, or systemic embolism and who had no more than two risk factors can be enforced by random sampling after the cohort is formed (repeat sampling with appropriate adjustment of the sampling variance estimator can be used to ensure that no information is discarded). – Pregnant women will be excluded. – Criteria related to planned treatments (e.g., "planned cardioversion") will not be reflected in the routinely collected data (because intent cannot be determined retrospectively); instead, we will take the initiation of the treatments of interest to imply that the criterion was not met. Similar certain exclusion criteria will also not need to be reflected because the initiation of treatment is indicative of them not being present (e.g., in clinical practice anticoagulation therapy is not started in the presence of active internal bleeding). – Platelet count criteria will not be used in the claims analyses because of expected high degree of missingness. We might use them in EMR-based analyses. – Aspirin-related criteria will not be implemented because claims (and EHR) data is highly incomplete for over-the-counter medications. Again, we will take treatment initiation as implying that co-treatments were appropriately chosen. – Other medication exclusions will be implemented using pharmacy claims. – Creatinine clearance will also not be used as an eligibility criterion in the main analyses. In our prior work the missingness in serum creatinine values was fairly high (more than 30% even among subsets with diagnostic or procedure codes related to chronic kidney disease). – The exclusion criterion related to serious illness with limited life expectancy will be operationalized by excluding individuals with severe illness at baseline (e.g., advanced stage chronic kidney disease, metastatic non-melanoma skin cancer, etc.)

Treatment strategies	Rivaroxaban (20 mg daily or 15 mg daily in patients with a creatinine clearance of 30 to 49 ml per minute) or adjusted-dose warfarin (target international normalized ratio [INR], 2.0 to 3.0). Patients in each group also received a placebo tab- let in order to maintain blinding.	– Treatments will be identified using pharmacy claims. The study design will be emulated using the methods described by Hernan and Robins (in simple parallel arm trial designs, the methods are equivalent to the standard new-user designs).[1,2,4] – It is not possible to emulate placebo control using routinely collected data.
Assignment procedures	We will emulate baseline randomization via different analytical strategies, all of which rely on assumptions of conditional exchangeability of the treated and untreated individuals, conditional on baseline covariates. Please see the **Section 11** for a summary of the different methods in our analysis plan.	
Outcomes	Stroke was defined as a sudden focal neurologic deficit of presumed cerebrovascular etiology that persisted beyond 24 hours and was not due to another identifiable cause. An event matching this definition but lasting less than 24 hours was considered to be a transient ischemic attack. Brain imaging (computed tomography or magnetic resonance imaging) was recommended for all suspected strokes, and this was performed in 82.1% of patients with strokes.	Stokes and transient ischemic attacks will be identified as for the eligibility criteria, using established diagnostic algorithms (e.g., as for our prior work using OptumLabs data), combining diagnosis codes and medication history information.
Followup procedures	The protocol required that all randomized patients be seen at 1, 2, and 4 weeks and monthly thereafter for the duration of the study for measurement of international normalized ratio (INR), surveillance for primary endpoint events, transient ischemic attack, myocardial infarction, medical or surgical procedures, adverse events, and vital status. A standardized questionnaire and examination were used to screen for stroke symptoms and potential clinical events during follow-up. Concomitant use of aspirin up to 100 mg daily was permitted. Treatment with thienopyridine antiplatelet agents was prohibited for 5 days before randomization and throughout the treatment period, except for patients undergoing cardiovascular interventions who were eligible to receive appropriate dual antiplatelet therapy with aspirin and a thienopyridine concomitantly with the assigned anticoagulant at the investigator's discretion.	– It is not possible to operationalize the methods for monitoring and increasing adherence or reducing loss to followup that were used in the trial. Instead, we will perform various adherence- and censoring-adjusted analyses. Please see **Section 11** for details. The median followup in the ROCKET trial was approximately 2 years and the longest reported followup up time was approximately 2.3 years.[7] Thus, it seems reasonable to use 2.5 years from time zero as the end of followup for this emulation. In stability analyses we will use the entire available followup but still focus our comparisons across methods and between the trial and the emulations on the 3-year followup duration.
Target parameters (causal contrasts)	We will estimate a number of different target parameters, all of which would be reasonable summaries of the treatment effectiveness in the randomized trial. Please see **Section 11** for details.	
Analysis plan	Please see **Section 11** for details.	

Table 2: Main aspects LEAD-2 trial emulation.[9]

Component of emulation	LEAD-2 trial emulation (criteria abstracted from the trial protocol)	Notes for the operationalization of the eligibility criteria in OptumLabs (see also our remarks regarding variable definitions in Section 3b)
Eligibility criteria	– Type 2 diabetes were screened and enrolled if they were – 18 – 80 years of age, – A1C between 7 and 11% (prestudy OAD monotherapy for □3 months) or between 7 and 10% (prestudy combination OAD therapy for >=3 months), and had BMI <=40 kg/m^2. – *Exclusion criteria:* insulin during the previous 3 months (except short-term treatment).	– Age criteria will be implemented using the corresponding claims or EHR field. – Diabetes will be coded using established diagnostic algorithms (e.g., as for our prior work using OptumLabs data), combining diagnosis codes and medication history information. – Medication exclusion criteria will be implemented using pharmacy claims. – BMI information may be used in the EMR data if available and adequately complete.
Treatment strategies	One of three once-daily doses of liraglutide (0.6, 1.2, or 1.8 mg/day) injected subcutaneously in combination with metformin vs. liraglutide placebo with metformin monotherapy (placebo group) vs. combination therapy with glimepiride and metformin (4 mg glimepiride once daily with the first meal of the day).	– We will examine whether the data permit a reliable identification of the different drug dosages using pharmacy claims information. We anticipate that this will be even harder in the EHR data. – It is not possible to emulate placebo control, the double dummy design, or the recommendation to take medication at a specific time of day using routinely collected data.
Assignment procedures	Randomization occurred after a 3-week forced metformin titration period followed by a 3-week metformin maintenance period. Subjects taking metformin at enrollment could go through a modified titration period or advance directly to the metformin maintenance period. After randomization, subjects underwent a 2- and 3-week titration period for liraglutide (up to 0.6, 1.2, or 1.8 mg, as per randomization, at 0.6-mg increases per week) and glimepiride (up to 4 mg, with 1-, 2-, and 4-mg doses at weeks 1, 2, and 3). Glimepiride (active and placebo) was taken orally once daily in the morning. Liraglutide was injected subcutaneously once daily at any time of the day in the upper arm, abdomen, or thigh using a pen injector device. Subjects were encouraged to inject liraglutide at the same time each day. The titration period was followed by a 23- or 24-week maintenance period during which the doses of study drugs were to be maintained. Metformin could be decreased to a minimum of 1,500 mg/day in the case of hypoglycemia or other adverse events but had to be maintained between 1,500 and 2000 mg/day during maintenance.	– The titration and run in period design can be (approximately emulated) using the methods for time-varying treatments described previously by Hernan and Robins.[4] – The dose adjustments will be addressed in adherence-adjusted analyses; please see **Section 11** for details. – We will emulate baseline randomization via different analytical strategies, all of which rely on assumptions of conditional exchangeability of the treated and untreated individuals, conditional on baseline covariates. Please see the **Section 11** for a summary of the different methods in our analysis plan.
Outcomes	The primary outcome measure was change in A1C at the end of the study.	We expect fairly severe missingness for this outcome in the claims (linked with lab results) and the EHR data. As recommended in the RFP we will use A1C data both from claims and EHR data. Please see **Section 11** for details of our approach to missing data.
Followup procedures	26 weeks post-time zero (randomization)	– We will emulate the study duration using a reasonable window for measurements.
Target parameters (causal contrasts)	We will estimate a number of different target parameters, all of which would be reasonable summaries of the treatment effectiveness in the randomized trial. Please see **Section 11** for details.	
Analysis plan	Please see **Section 11** for details.	

b. Variables

A note about variable definitions: In the sections below, we describe our general approach to variable definitions and coding for the OptumLabs claims and EMR data. In our view, the process of operationalizing the variables into code algorithms (e.g., selecting the codes to be used, or deciding the positions from which codes will be extracted for covariates or outcomes from claims data) should be part of the actual project, not carried out at the proposal step, to ensure that all decisions (and processes for making those decision) can be pre-registered with the initial version of the study protocol (which will be generated, based on this proposal, at the study start). That way, any changes or deviations can be prospectively documented before being implemented in the data. Specifically, within the first week of starting the project we will conduct two separate (one for the ROCKET trial emulation and one for the LEAD-2 trial emulation) focused mini-reviews to identify candidate algorithms (preferably, validated) for claims and EHR data, for information needed to define the eligibility criteria, other baseline covariates, and outcomes in the two emulations. We note that is it easy to obtain several examples of variable definition choices and code algorithms by established research teams, specifically using Optum data, in settings related to the ROCKET trial emulation[10-15] and the LEAD-2 trial emulation.[9,16] Our team has experience implementing such algorithms in routinely collected data (e.g., private insurance, Medicare) and specifically in the OptumLabs data (for claims, laboratory test results linked with claims, and EHR data).

Key Outcomes, if applicable

Our understanding based on the RFP is that only primary outcomes are to be emulated. Please see **Tables 1** and **2** for more information about the primary outcomes in each emulated trial. We note that the distinction between primary and secondary outcomes in observational studies is much less important and that several of the secondary outcomes of the trials can be operationalized in claims data using validated algorithms. As noted in the RFP, for the LEAD-2 trial emulation, we expect that additional HbA1c values can be obtained from the EHR for patients missing HbA1c values in the claims data. This will expand the number of patients with HbA1c outcome data beyond the HbA1c results found in the claims files alone.

Key Indicator(s) of Interest, if applicable

The “key indicators” will be the randomized interventions in each emulation (see **Tables 1** and **2** for details of the treatment strategies and **Section 11** for more information about our analysis plan for handling treatment non-adherence).

Demographics, Patient Characteristics, and Other variables

We expect that we will use a large number of variables for confounding control, adherence- and censoring-adjusted analyses, or to explore heterogeneity of treatment effects (following any prespecified subgroup analyses reported in the trial). Our variable definitions will use diagnostic and procedure codes from claims or variables extracted from EHR data, as needed. In a recent OptumLabs project on chronic kidney disease we coded more than 200 baseline covariates and we expect a similar or larger number of covariates to be collected for this project. As noted in the RFP and expected on the basis of our prior OptumLabs work, for both the ROCKET and LEAD-2 emulations, we expect that the subset of patients with linked EHR data will have

additional covariate information (e.g., blood pressure, BMI, and various lipid measurements) which we plan to use in analyses.

Other Variables, if applicable

Not applicable.

c. Study limitations

The major limitation in studies using routinely collected observational data to emulate target trials is *confounding by unmeasured (unobserved) variables*. The extent of confounding due to unmeasured variables (1) cannot be identified from the observed data alone; (2) can have arbitrary direction and magnitude. Because of these properties, it remains a threat to validity despite the use of high-quality data and state-of-the-science causal and statistical methods. The main approach for addressing such confounding is to conduct sensitivity analysis to examine the impact of confounding by unmeasured variables on the study results. In previous work using OptumLabs data, we have implemented non-parametric sensitivity analysis methods and we will draw on that work if needed for the current project. We do note, however, that in the context of sensitivity analyses, an observational study does not produce a single summary measure of any of the target parameters listed in **Table 3**; instead, we can only obtain a curve of treatment effect estimates and corresponding confidence limits, over values of unidentifiable sensitivity parameters. Furthermore, the degree of sensitivity to confounding by unmeasured variables is parameter and estimator dependent. Nevertheless, methods for sensitivity analysis may be useful for contextualizing the results of this project, especially if the observational analyses disagree with the results of the randomized trials.

We also recognize additional limitations related to *data quality issues, such as high dropout (especially incomplete followup in EHR data), covariate missingness, or incomplete linkage between claims and EHR data.* As noted in **Section 11**, some of these issues will be mitigated by our proposed analysis methods (e.g., loss to followup via the use of inverse probability of censoring methods). Furthermore, these limitations may be indirectly addressed by the triangulation possibilities engendered by the use of different cohorts (claims or claims linked with EHR data, with or without additional covariates).

3. Translation Potential

Translation sponsors or champions at the School of Public Health and the Brown University at large (i.e., the institution with which Dahabreh has his primary affiliation):

Name	Title	
Karen Scanlan	Director of communications and outreach, School of Public Health, Brown University	
Brian Clark	Director News and Editorial Development, Brown University	

a. Consumers/ Patients

Although the goals of our project are methodological, we will contribute to the clinical literature on major chronic diseases (e.g., diabetes, atrial fibrillation). Thus, both the emulation results and the expanded population results should be of direct interest to patients and carers

b. Payers

The project will generate results that can benefit payers using observational data analyses to inform their decision-making. CESH has established connections with payers such as BCBSRI and Medicaid. After completion of the project we plan to share our summary findings (e.g., published papers) with these partners and engage them in discussions about the methodological and substantive implications of our findings with their work.

c. Providers
d. Life Science Companies
e. Employers
f. NGOs
g. Medical societies
h. State or federal agencies/programs

We address items **c.** through **g.**, listed above, collectively. Providers, life sciences companies, employers, NGOs and Medical Societies stand to benefit from the results that will be produced from this project. To give some examples, many NGOs are interested in sponsoring randomized trials of interventions and in developing policies based on the results of these trials. Projects such as this (and especially the larger followup study mentioned in the RFP), if they find that observational studies designed to emulate target trials can approximate the results of the actual trials fairly accurately and reliably, would suggest that NGOs can make their research more efficient (by conducting observational studies) and base their policy recommendations on a broader evidence base (by admitting to the evidence base well-conducted observational studies). A similar argument applies to medical societies (e.g., when issuing clinical practice guidelines) or life sciences companies (e.g., when determining whether development or marketing decisions can be informed by well-conducted observational studies, in addition to trials). Last, as noted in the RFP, trial emulations may offer a cost-effective way of examining label expansion possibilities and monitoring safety.

i. Your organization

The Center for Evidence Synthesis in Health at Brown University and the Causal Inference Program at the Harvard School of Public are committed to the conduct and dissemination of methodological research on causal inference methods, including methods to improve the design and analysis of observational studies using routinely collected data.

j. Other

Not applicable.

4. Publication and Dissemination of Results Checklist

List the communication channels most appropriate for reaching target audiences with key findings.

Please see Section 17 of this document for OptumLabs publication review requirements.

☒ Peer-review/professional journal articles; expected article(s), lead author, other authors and target journal(s):

We expect at least two peer reviewed publications, to be submitted to clinical journals, and one methodological publication, with the following provisional titles:

1. Emulating a target trial comparing rivaroxaban vs warfarin for patients with atrial fibrillation using routinely collected observational data
2. Emulating a target trial comparing of pharmacotherapies for patients with diabetes using routinely collected observational data
3. Target parameters, causal models, and estimation methods in two emulations of randomized trials using large, rich routinely collected data

We expect that manuscripts from this project will be co-authored by all team members, and led/senior authored by Dahabreh or Hernan.

☒ Professional meeting presentations/posters/abstracts; targeted meetings:

1. Society for Epidemiologic Research; expected deadline for submission: January 2020; expected presentation date: in the Summer of 2020
2. ICHPS; expected deadline for submission: May 2019 (for preliminary results, + 1 year for final results); expected presentation date: January 2020 (for preliminary results, + 1 year for final results)
3. Atlantic Causal Inference Meeting; expected deadline for submission: January 2020; expected presentation date: Spring of 2020
4. American Heart Association annual meeting; expected deadline for submission: Spring-Summer 2019 (for preliminary results, + 1 year for final results); expected presentation date: November 2020 (for preliminary results, + 1 year for final results)
5. American Diabetes Association Annual Meeting: expected deadline for submission: Fall-December 2019; expected presentation date: June 2020

In addition to submitting abstracts for presentation at the above meetings, if our proposal is selected for funding, we will invite the second team independently attempting the same emulations, *to co-organize sessions/ workshops in the above scientific meetings to disseminate the work, in collaboration with OptumLabs and external discussants.*

We note that travel to the above listed meetings will be covered from institutional resources (thus it has not been budgeted).

☒ Press release / media coverage

☒ Media packages

☒ Social media distribution

☒ Policy briefs
☒ White paper
☒ Professional, medical and advocacy group communication: e.g., professional societies, NGOs, life sciences companies
☐ Shared patient and/or provider decision aides
☐ Video clips / multimedia
☒ Articles or blogs targeting specific stakeholders
☒ Other web posts/content
☐ Patent applications
☐ Investor packages
☐ Translation to practice and/or commercialization plan
☐ Other: Not applicable.

5. Resource and Timeline Planning

Research Project Milestones	Description of Milestones	Planned Date
Research Application Submission*	Applicant submits the complete Research Application to OptumLabs for review.	January 2019
SOW Completion	OptumLabs and your institution have signed an SOW for the research project. Please contact your Partner Relationship Manager with questions about the SOW process and timeline.	Consultation with the Partner Relationship Manager to occur in January-February 2019
Project Initiation*	Project planning and provisioning complete (typically when vDI is provisioned to team, approx. 10-15 days after SOW signed).	April 1, 2019
Data Set Completion	For some projects, this may be very fast, i.e. if a standard view is used and population definition is straightforward. Some projects will require significantly more effort to define the data set, or may have contingencies based on project specific circumstances.	May 30, 2019
Analysis Complete	Completion of all statistical and descriptive analysis. Study is ready to move into reporting phase.	September 2019 (emulation); December 2019 (expanded population)
Submit Report and/or Manuscript Draft for OL Review**	OptumLabs must review all research materials and communications 30 calendar days prior to external submission to ensure they do not violate any legal, compliance, and/or privacy policies.	September 2019 (emulation) December 2019 (expanded population)
Report/Manuscript Complete*	Completion/issuance of the first document that includes analysis and interpretation. This could be a study report, manuscript, abstract, poster or other form of communication. If it is for publication or presentation, this should be the date submitted. *Note: It is recognized that multiple reports may come out of the same project. Depending on the work involved, subsequent reports could become new projects. Some projects may not result in a full report. In this case, a simple closeout report or memo can be issued.*	December 2019 (at least 2 reports, one for each trial)
Next Steps	Discuss opportunities for translational and/or commercialization transfer based on results of the study	December 2019 (and throughout the project)
Project Completion or Termination*	Final expenditure of resources completed	December 2019

(*) Indicates that dates are required

(**) See **Section 17** for OptumLabs publication review requirements

Please specify if the project start is relative to another event (e.g., SOW signature, one month after receipt of grant, etc.):
The project will commence upon funding announcement. We expect this to occur in February 2019.

6. Resource, Staffing, and/or Collaboration Plan

a. Staffing and Collaboration

Role	Name	Organization	Covered by confidentiality terms with OptumLabs (yes/no)	Duration (weeks)	Time (hrs/week)
Principal Investigator	Issa Dahabreh	Brown University	Yes	~44	6 hours per week
Senior consultant	Miguel A. Hernan	Harvard University	No, to be instituted while awaiting the funding decision	~44	2 hours per week
Analyst	Hongseok Kim	Brown University	Yes	~44	10 hours per week
Analyst (graduate student)	Sarah Robertson	Brown University	Yes	~44	20 hours per week
Programmer	TBN	Brown University	Yes	~44	9 hours per week (average)

b. Project Team Capabilities

Issa Dahabreh, MD MS, Assistant Professor of Health Services, Policy and Practice and founding member of the Center for Evidence Synthesis in Health (CESH), will serve as PI for this project. He serves as Associate Director of the AHRQ-designated Brown University Evidence-based Practice Center (EPC), one of 13 such centers in North America. His research interests include the evaluation of methods for drawing causal inferences from observational data, generalizing the results of randomized trials to new target populations (in which no experiments can be conducted), and synthesizing evidence from diverse sources. He teaches the 2-semester PhD-level sequence on Methods for Health Services Research. He has led and collaborated on numerous projects involving the analysis of observational data of varying complexity, including two PCORI Methods Research Awards (see **Section 7d**), one of which used OptumLabs data (among other large sources of observational data) to emulate target trials in cardiovascular and kidney disease. To our knowledge, that project was one of the first OptumLabs projects to use, in addition to the claims data, laboratory test results that have been linked with the claims, and to combine claims and EHR data to obtain additional baseline covariates for confounding control.

Miguel Hernan, ScD, Kolokotrones Professor of Biostatistics and Epidemiology at the T.H. Chan – Harvard School of Public Health, will be Senior Consultant on this project. His research is focused on learning what works for the treatment and prevention of diseases like cancer, cardiovascular disease, and HIV infection. Together with collaborators, he has designed numerous analyses of healthcare databases, epidemiologic studies, and randomized trials. For over a decade, he has coordinated the HIV-CAUSAL Collaboration, a multinational consortium of prospective studies from Europe and the Americas. Hernan is internationally recognized for his methodological work on the design and analysis of observational studies emulating target trials, structural treatment of biases, and instrumental variable analyses. We note that Hernan's formal role as a consultant was chosen for budgetary and administrative purposes (to facilitate the timely execution of the project, if funded); in terms of his role, he will be co-leading the project with Dahabreh.

Iman Saeed, MSc, Research Associate, is a Biostatistician at CESH. She has worked with Dahabreh on a number of projects using causal inference and missing data methods during her MSc training (Biostatistics, Brown University, 2017). Currently she is working with Dahabreh, Hernan, and collaborators on sensitivity analysis methods for studies combining a randomized trial with observational data to extend causal inference from trial participants to the target population with funding from PCORI. She will work on back-end analytical methods, supporting Robertson and the TBN programmer who will have hands on access to the OptumLabs data.

Sarah Robertson, MSc, PhD candidate, Research Associate, is a Biostatistician at CESH who is also pursuing a PhD degree in the Department of Health Services Policy & Practice, with Dahabreh as her advisor. For the last 4 years she has worked with Dahabreh and Hernan during her MSc training (Biostatistics, Brown University, 2016) and as staff at CESH, on projects related to the use of observational data to emulate target trials (including experience with OptumLabs data) and methods for extending inferences from randomized trials to a new target population.

MSc-level Programmer, TBN. The to-be-named programmer will be selected among the many expert programmers in the Department of Health Services Policy & Practice, at Brown University. All programmers have extensive experience using routinely collected data.

Relationships among team members and summary of joint work on projects related to this proposal: Dahabreh and Hernan (with colleagues, including Robertson and Saeed), during the last 4 years, have collaborated on developing a theory (i.e., a collection of concepts and identification, estimation, and sensitivity analysis results) of extending randomized trials to new target populations using observational data.[17,18] Furthermore, Dahabreh and Hernan have worked together on several projects emulating target trials using observational data, including projects using OptumLabs data (claims and EHR), in chronic kidney and cardiovascular disease with funding from PCORI (Methods Research Award ME-1306-03758, PI: Dahabreh). Lastly, they are currently collaborating on a systematic attempt to systematically evaluate whether SEER-Medicare data can be used to emulate target trials in Oncology. Dahabreh has long been interested in formally comparing treatment effect estimates from observational studies against those of randomized trials addressing the same clinical question. He has published two such comparisons using data extracted from published sources,[19,20] including the first comparison of this kind that focused on propensity score methods.[19] In more recent work, he has identified the limitations of comparisons based on published data, has advocated for research of the type described in this RFP, and has undertaken such research with funding from PCORI.[21,22] Hernan has been a leader in advocating for the conceptualization of observational studies as attempts to emulate a well-specified target trials, in general for comparative effectiveness research and in particular for observational studies using routinely collected data[3,4,23-28] (the broad idea that observational studies should aspire to an “experimental ideal” has a much longer history, possibly dating as far back as 1944).[29]

c. References

Please see the reference list at the end of the document, **Section 14**. We have taken the liberty of adding this section because the extensive reference list would be disruptive if presented at the middle of the document.

d. List other relevant work that your institution has handled previously or is engaged in currently

We only list the three most recent and most relevant project on which Dahabreh and Hernan, the senior personnel on this proposal, have collaborated or are actively collaborating on. The much more extensive list of relevant work by our team members is better, but still not fully, represented in the references.

- **PCORI Methods Research Award ME-1306-03758**, PI: Dahabreh (final report completed in 2018). This 3-year study included a number of attempts to emulate target trials using observational data, including several that used OptumLabs data.
- **PCORI Methods Research Award ME-1502–27794**, PI: Dahabreh (ongoing). This study is developing new methods for studies combining randomized trials and observational data (including routinely collected data) to extrapolate causal inferences from the trial population to the target population represented by the observational data.
- **U.S. National Cancer Institute contract HHSN261201700047**, PI: Hernan (ongoing). This study is using SEER-Medicare data to emulate four target trials in oncology.

e. Software and programming resources

Please select software/programming language(s) that may be used for data set creation and/or analyses of patient-level data. Software used for these purposes must be installed in the OptumLabs environment. Aqua Data Studio (SQL querying tool) and R/RStudio are provided.

For data extraction and data set creation (i.e. creating a patient-level data set using the OLDW):

DBVisualizer (SQL)
SAS

If other, please specify and describe use:
No other software will be needed.

For analyses of patient-level files:

R and RStudio
Stata
SAS

If other, please specify and describe use:
No other software will be needed. Note that we will use our institutional SAS license and provide new Stata licenses for our analysts.

c. Non-standard software to be provisioned to vDI

Please provide specific details below. Note: Applicant is responsible for contracting and licensing software from vendor or OptumLabs if available.

No non-standard software will be needed.

Software (Name/Manufacturer)	Version #	Name/Email of install contact
Not applicable.		

7. Special Project Related Approvals

a. IRB Approval

Will this project require IRB approval from your institution? No, please see below for details.
If yes, please estimate duration required for IRB approval (if known)

At Brown University, projects like this require consultation with the IRB and an attestation by the Principal Investigator that the data are de-identified. Once this is provided, the IRB furnishes a formal letter waiving the need for further review. In that last three projects led by Dahabreh, including one that used OptumLabs data, this process has taken less than 2 weeks to complete.

b. Publication Approval

Does any party outside your institution have rights to review, approve, or control dissemination of the reports of this project? Please explain.

No outside part has rights to review, approve, or control dissemination of the reports of this project.

c. External Data Approval

Will this project include any external data being brought into the project sandbox, beyond simple code lists?

No.

If yes, this file will need to be reviewed and approved for compliance by OptumLabs. Has this file been reviewed and approved?

Not applicable.

A description of the file will be needed for the project SOW.

Not applicable.

8. Services Requested from OptumLabs

a. Staff Services from OptumLabs

None requested.

Resource Type	Deliverables	Estimated hours	Estimated service dates
Not applicable.			

b. Other Services from OptumLabs

None requested.

Service Description*	Date needed	Comments
Not applicable.		

9. Analysis Plan

a. Methods

A note on the choice of analysis methods: For analysis results to have a tenable interpretation, it is important that the analysis methods be appropriate for the target causal parameters of interest. Thus, when comparing estimates from a collection of specific statistical methods, care is needed to ensure that all the methods under consideration can be used to estimate the same causal parameter (and that they can do so under compatible causal assumptions). Consider two concrete examples of direct relevance to this proposal: (1): ***single equation, multivariable single equation regression methods*** (i.e., maximum likelihood estimation of the coefficients of a multi-variable regression) estimate conditional average treatment effects (conditional on all the covariates entered in the regression together with the treatment indicators). For ***non-linear regression models*** (including the popular logistic regression and Cox proportional hazards regression models) these ***conditional average treatment effects are not in general equal to the corresponding population-averaged (marginal) causal effects.*** For instance, the coefficient from a multivariable logistic regression model of the outcome on treatment and covariates does not equal to the marginal odds ratio, *even if* the multivariable regression model is correctly specified and in the absence of any unmeasured confounding (this phenomenon relates to the non-collapsibility of the odds ratio; in general the conditional odds ratio is expected to be further away from the null compared to the marginal odds ratio, except in rare cases). (2) ***Inverse probability weighting, matching, and instrumental variable estimators do not estimate the same population-averaged causal effects:*** Specifically, ***inverse probability weighing*** targets the average treatment effect in the entire population (or the average treatment effect in the treated/controls, with appropriate choice of odds weights). ***Matching methods*** estimate causal effects in populations that depend on the data generating mechanism and the matching algorithm (e.g., 1-to-1 matching without replacement, when one treatment is much less commonly prescribed than the comparator produces estimates of the average treatment effect in the population treated with the rarely prescribed treatment). Lastly, ***instrumental variable methods***[30,31] estimate different causal parameters depending on auxiliary assumptions that are untestable; under homogeneity, they estimate the average treatment effect; under certain structural nested mean models

assumptions, they estimate the average treatment effect on the treated; and under monitonicity, the estimate the so-called average treatment effect on the compliers.*

Choice of causal models and methods of estimation for the current project: With the above point in mind, in **Table 3**, we have identified target parameters that are of scientific and policy interest for the ROCKET trial and the LEAD-2 trial emulations (the same methods would apply to the corresponding "expanded population" analyses). The ***target parameters*** we have listed are identifiable under randomization (in the actual trials) and thus are natural choices for the emulation analyses. For each of these target parameters we have described ***causal (counterfactual) models*** under which the identification of the same target parameters can be considered when using observational data (in the emulations and the expanded population analyses). Last, for each target parameter – causal model combination we have listed our choice of ***estimation methods***. It is not feasible to give complete details for the implementation of each of these analysis methods within the scope of the current proposal (full details will be provided in the study protocol to be generated is this proposal is selected for funding).[4,32-36] Instead, we have provide here a number of citations to technical papers or book chapters that describe the implementation of the method in the table (we have intentionally focused on works by Hernan and collaborators to establish that we have the capability to implement the methods for the project's purposes). We now discuss aspects of the analyses that are cross-cutting, in the sense that they can influence the results of many of the listed techniques.

Censoring- and adherence-adjusted analyses: These two types of analyses, though different in their goals, are closely related because they involve the estimation of treatment effects of time-varying interventions (to prevent censoring over time; or to enforce a particular adherence level). For these analyses we expect to work in discrete time (we expect that 1 month intervals will be feasible and give a reasonable approximation to the underlying continuous functions) and define variables that capture treatment status based on prescribing information (e.g., prescriptions filled, number of pills). The discrete time structure is reasonable for claims data and allows for practical ways of collecting information on time-varying exposures and treatments. Note that the discrete time structure allows the computation of functions of cumulative exposure that can be made arbitrarily flexible and thus can (in theory) approximate the true cumulative exposure-outcome function, while properly adjusting for time varying covariates. We note that popular adherence measures such as medication possession ratios, when entered in conditional outcome models or models for the propensity score produce invalid results in the presence of time-varying confounding (except under highly implausible assumptions).[32,33] In adherence-adjusted analyses, we will use clinically meaningful definitions of adherence.[23] Specifically, individuals who discontinue treatment due to toxicity or other clinically mandated reasons will be considered to be adhering to the emulated trial protocol, as no reasonable protocol would force patients to stay on treatment in these situations. For censoring due to death, we will use the suggestions by Young et al.[37] to conceptualize the target parameters and refine our analysis choices.

Modeling and robustness to model misspecification: In addition to the causal models described in **Table 3**, our analytical methods require the specification of working models for a number of regression functions (conditional probability or expectation functions), such as

* The difficulties with instrumental variables outlined here, together with our belief that useful and valid instrumental variables are exceedingly hard to find in routinely collected claims or EHR data, are the reason why we decided not to include instrumental variable methods in this proposal.

models for the probability of treatment, the outcome, censoring, and treatment adherence over time. Our main analyses will use ***parametric working models*** because they are the most commonly used approach in practice and because in our experience – if made sufficiently flexible – they approximate the results of semiparametric or nonparametric models fairly closely. As a stability analysis, we will repeat our analyses using semi-parametric methods to estimate the working models. Specifically, we will use random forest methods (when modeling discrete responses), and lasso or elastic-net regularized generalized linear models (for both binary and continuous responses). Because we plan to work in discrete time for survival analyses, and in censoring- and non-adherence adjusted analyses, the corresponding models will be pooled (over time periods) generalized linear models for binary responses.

When using parametric models, our initial modeling strategy ***will be to enter all variables deemed likely confounders in the models on the basis of substantive knowledge***,[38] with flexible terms for continuous variables (restricted cubic splines with 5 knots at the default positions) and product terms between variables when indicated by substantive knowledge or model specification assessments (e.g., covariate balance after inverse probability weighting or after propensity score matching). Note that in the stability analyses using semi-parametric models sparsity-appropriate model selection is built into our chosen procedures.

We will model time nonparametrically (e.g., with appropriate indicator variables in discrete time), as needed (e.g., in survival analyses or adherence- and censoring-adjusted analyses). When nonparametric modeling of time is not possible (e.g., due to rare events) we will begin by using restricted cubic splines with 7 knots and then progressively reduce the number of knots until the models can be estimated.

Whenever the data allow, we will not make homogeneity assumptions among treatment groups (e.g., outcome regression models and probability of censoring models will be fit separately by treatment group).

To reduce the impact of possible model misspecification, when available, we have chosen to include multiply robust estimators (both estimating equation-[39] and TMLE-based[40,41]) in our analysis plan (Table 3). These estimators utilize multiple working models and remain consistent and asymptotically normal when either model (but not necessarily both) is correctly specified. For technical reasons that are beyond the scope of this proposal, we will only use semi-parametric methods to estimate working models in combination with doubly robust estimators.[39,42]

Table 3: Target parameters, measures of treatment effect, causal models, and methods of estimation.

Target parameter	Measure of effect	Causal models; methods of estimation	Notes about additional assumptions or special issues in estimation
ROCKET TRIAL (failure time outcome)			
Intention-to-treat conditional treatment effect	Conditional hazard ratio (averaged over the follow-up period if the conditional counterfactual hazards are not proportional)	(1) Multivariable adjusted Cox regression model; maximum partial likelihood	These approaches correspond to the "single equation method" asked for in the RFP. Note that marginalizing the estimated Cox model over the baseline covariates is equivalent to approach (2) below, except for the assumption of conditional proportional hazards.
Intention-to-treat effect in the total population of eligible individuals (unadjusted for non-adherence)	Marginal difference of survival probabilities as specified time points & survival curves for each treatment	(2) g-formula computation; outcome model-based standardization following outcome regression (3) Marginal structural survival probability model with time-varying marginal counterfactual hazards; inverse probability of treatment weighting (4) Efficient and doubly robust estimation using estimating equations (5) Efficient and doubly robust estimation using TMLE	All approaches listed here do not require proportional hazards assumptions over the treatment indicator. Approach (2) adjusts both for confounding and censoring. Approaches (3) through (5) require combination with inverse probability of censoring weighting to adjust for censoring. We will perform both censoring adjusted and unadjusted analyses. Approaches (4) and (5) are based on the efficient influence function and are locally efficient and doubly robust.
	Marginal survival ratio comparing treatments & marginal hazard ratio (using a Weibull distribution)	(6) Nested structural accelerated failure time model; g-estimation	When using a Weibull parametric family, we can recover a hazard ratio from the accelerated failure time model, using the dual interpretation of Weibull regression as a proportional hazards and accelerated failure time model.
	Marginal hazard ratio (averaged over the follow-up period if the marginal counterfactual hazards are not proportional)	(7) Marginal structural Cox proportional hazards model; inverse probability of treatment weighting	Assumes proportional marginal counterfactual hazards among treatment groups.
Intention-to-treat effect in the treated (unadjusted for non-adherence)	Marginal difference of survival probabilities as specified time points & survival curves for each treatment	(8) Matching on the estimated probability of treatment, followed by non-parametric modeling of the survival probability as a function of time, in each treatment group (9) Marginal structural survival probability model with time-varying marginal counterfactual hazards (in the total study sample); inverse *odds* of treatment weighting	Using indicator functions for time-periods and interactions with treatment group this approach avoids proportional hazards assumptions. Naïve nonparametric bootstrap inference is not appropriate for some matching estimators because they are non-smooth (see the text of the proposal for how we intend to address this).
	Marginal hazard ratio	(10) Matching followed by Cox proportional hazards	Same as above.

	(averaged over the follow-up period)	regression (11) Marginal structural Cox proportional hazards model; inverse probability of treatment weighting	
Per-protocol effects (i.e., adjusted for non-adherence)	Marginal difference of survival probabilities as specified time points	(12) g-formula computation; outcome model-based standardization following nested regression (13) Marginal structural model with time-varying hazard; inverse probability of treatment weighting (14) g-estimation of accelerated failure time model	These methods can account for observed non-adherence and estimate the effects of complete adherence or partial adherence.
LEAD-2 trial (continuous outcome evaluated at the end of the study)			
Intention-to-treat conditional treatment effect	Conditional mean difference ratio	(15) Multivariable conditional outcome model; adjusted ordinary least squares regression	These approaches correspond to the "single equation method" asked for in the RFP. Note that marginalizing over the baseline covariates is equivalent to approach (13) below.
Intention-to-treat effect in the total population of eligible individuals (i.e., unadjusted for non-adherence)	Marginal mean difference at the end of the study	(16) g-formula computation; outcome model-based standardization following outcome regression (17) Marginal structural mean model; inverse probability of treatment weighting (18) Efficient and doubly robust estimation using estimating equations (19) Efficient and doubly robust estimation using TMLE (20) Nested structural mean model; g-estimation	Approach (16) also addresses early dropout from the study. Approach (17) needs to be combined with inverse probability of outcome missingness (i.e., no early drop-out) to account for individuals who do not have measurements at the end of the study. Inverse probability of outcome missingness weights can also be used with approaches (18) through (20).
Intention-to-treat effect in the treated (i.e., unadjusted for non-adherence)	Marginal mean difference at the end of the study	(21) Matching on the estimated probability of treatment (i.e., the propensity score) (22) Marginal structural mean model (in the total study sample); inverse *odds* of treatment weighting	Using indicator functions for time-periods and interactions with treatment group this approach avoids proportional hazards assumptions. Naïve nonparametric bootstrap inference is not appropriate for some matching estimators because they are non-smooth (see the text of the proposal for how we intend to address this).
Per-protocol effects (i.e., adjusted for non-adherence)	Marginal mean difference at the end of the study	(23) g-formula computation; outcome model-based standardization following nested regression (24) Marginal structural mean model for time-varying treatments; inverse probability of treatment weighting (25) Nested structural mean model for time-varying treatments; g-estimation	These methods can account for observed non-adherence and estimate the effects of complete adherence or partial adherence.

RFP = request for proposals; TMLE = targeted minimum loss estimation.

Dealing with extreme weights in weighted analyses: Many of the estimation approaches we propose to use involve the estimation of inverse probability weights (for treatment initiation, censoring, or non-adherence). These estimation approaches are sensitive to near-violations and violations of related positivity assumptions (i.e., the assumptions of positive probability of treatment assignment, of no-censoring, and of full adherence). We will address issues related to extreme weights using an explicit bias-variance tradeoff framework.[43] After estimating the weights, and before estimating treatment effects, we will begin by finetuning our selection criteria and examining the covariates that may be responsible for the extreme weights. If difficulties persist, we will repeat our analyses after trimming or truncating extreme weights using cut-offs ranging from the 99th to the 95th percentile of the weight distribution to define "extreme" weights.

Matching estimators: We have noted above the difficulties when comparing the results from matching estimators against other methods of estimation for population-averaged causal effects (e.g., IP weighting). In the table, we have listed matching as estimating the average treatment effect on the treated because our background knowledge and the preliminary counts reported in **Section 11b** suggest that the treatment group corresponding to the "experimental intervention" in the emulated trial will be less common than the treatment group corresponding to the "control intervention"; in such cases, matching methods (unless combined with additional weighting) approximately estimate the average treatment effect in the population of patients treated with the experimental intervention (i.e., the average treatment effect on the treated). We will use 1-to-many matching on the propensity score with replacement, within calipers of 0.1 standard deviation (on the logit-propensity score scale). Because matching estimators are non-smooth[44] we will consider to use of the m-out-of-n bootstrap[45] and the methods described by Otsu et al.[46] to improve the performance of bootstrap methods (Dahabreh is currently supervising a PhD student, not budget for in this proposal, who has already implemented methods to address these very issues).[46,47]

Issues of statistical inference and bootstrap methods: In general, our preference will be to provide inference (statistical tests and 95% confidence intervals) via the non-parametric bootstrap. Specifically, we will (1) sample the observed study data on covariates, treatments, and outcomes with replacement; (2) repeatedly perform the desired analyses collecting the results; and (3) examine the bootstrap distribution and use it for inference. In most cases and in large samples, we have found that the percentile bootstrap intervals and the normal-approximation Wald-style bootstrap intervals work reasonably well for the proposed statistical methods (and produce similar results). We will address any issues by using bootstrap variants that have better performance than the standard non-parametric bootstrap in challenging cases (e.g., as described for non-smooth matching estimators, described above). We will use the bootstrap results for inference on the counterfactual survival probabilities (in the ROCKET trial emulation), counterfactual outcome means (in the LEAD-2 trial emulation), and the treatment effect measures listed in **Table 3**.

Missing covariate and outcome data: We do not expect substantial covariate missingness in the claims data. Outcome missingness in claims analyses will be addressed by the censoring-adjusted analyses described above. We expect much greater outcome missingness when using laboratory results (linked to the claims or from the EHR), and also that HbA1c values will be obtainable from 2 separate sources (laboratory results linked with the claims data or via the EHR linkage to be performed for this project). Furthermore, when using laboratory results as regressors in working models, regardless of the data source, baseline and time-varying covariate missingness is likely to be a major issue in the both trial emulations. Our main analysis will use complete cases, in order to avoid the need for additional modeling assumptions

(e.g., when modeling the missing data conditional on the observed data under missing at random assumptions). As a stability analysis, we will repeat our analyses using multivariate normal multiple imputation.[48] In our prior OptumLabs work, we have found that multivariate normal multiple imputation performed well in this setting and returned results similar to multiple imputation using chained equations. To combine multiple imputation with the bootstrap we will use the approach suggested by Schomaker & Heumann.[49]

Comparisons between the observational analysis results and the target trial, and comparisons among different estimators: We will use the bootstrap results described above to compare treatment effect estimates from the observational analyses against the published randomized trial results. As suggested in the RFP, we will use the following criteria to organize our comparisons: (1) regulatory agreement, defined as statistically significant result with directional equivalence between the RCT and observational study; and (2) estimate agreement, defined as the point estimate of the observational study falling within the 95% confidence interval of the ATE from the RCT using the reported standard errors of the RCT to define the confidence interval. We also believe that the following additional comparison may be useful: (3) estimate the point estimate and confidence interval for the ratio or difference of the trial estimate vs. the corresponding observational study estimate (the choice between taking the difference of the estimators or their ratio will be based on whether the effect measure is additive or multiplicative, respectively).

Last, we will use the bootstrap results to compare the treatment effect estimators listed in **Table 3** when appropriate, that is, when the estimators share the same target parameter

b. Preliminary Patient Counts

Counts were obtained using OptumLab's Natural History of Disease tool using data current as of Jan 16, 2018.

Topic	Data years	
	1994 - 2018	From drug approval for indication - 2018
ROCKET trial emulation		
Individuals with atrial fibrillation claims	270,461	153,637
Individuals with at least one claim for warfarin treatment	102,089	48,380
Individuals with at least one claim for rivaroxaban	23,230	19,710
LEAD-2 trial emulation		
Individuals with diabetes claims	746,221	359,587
Individuals with diabetes and liraglutide claims	11,804	5,840
Individuals with diabetes and liraglutide + metformin claims (estimated value due to limitations of the NHB tool)	10,671	5,054

NHD = Natural History of Disease.

Although we expect substantial reduction in sample size as we apply the additional selection criteria listed in Tables 1 and 2, the above preliminary counts suggest that we will have adequate sample size to precisely estimate the target parameters of interest. Note that ***we have***

decided not to perform any power calculations for this proposal because of the undue influence they place on statistical significance (which is particularly inappropriate in observational studies) and because their meaning in the context of this project would be ill-defined (e.g., we cannot modify the sample size in response to any power calculation result and since the focus of the project is methodological any power calculation result would not be reflective of the true goals of the project).

c. Data to be imported to Project Sandbox, if applicable[2]

☒ Program codes: SAS, SQL, Stata, R, etc.
☒ Code lists: NDC, ICD-9, CPT, HCPCS, etc.

If additional data is required in the sandbox, please describe below:
No additional data will be imported.

d. Data to be exported from OLDW[1]

☒ Program codes: SAS, SQL, Stata, R, etc.
☒ Code lists: NDC, ICD-9, CPT, HCPCS, etc.
☒ Summary tables, charts and results (aggregated data)

e. AHA data fields

Not applicable.

10. Contingencies, Assumptions, and Risks

We believe there are two primary sources of risk for this project:

(1) Data-related issues that may limit our ability to emulate the target trials: This is meant to include all data-related issues that independent of our project management, data management, and ability to conduct statistical analyses may lead to failure to emulate one of the target trials, e.g., because of inadequate sample size, low event rate, or extreme data missingness. We think these are low probability events. First, and most important, there are several observational studies comparing the treatments that we plan to compare in the emulation and the expanded population analyses, suggesting that there is adequate sample size to perform the planned analyses (for example, there exist numerous observational studies of rivaroxaban or liraglutide using Optum or OptumLabs data).[9-16] The results reported in these studies, together with our preliminary patient count results reported in **Section 11b**, suggest that

[2] Data and analytic files cannot be directly imported or exported from the virtual desktop environment without specific permissions. Exporting of individual record data from the environment is not permitted.

the proposed analyses are feasible and will have adequate sample size. Second, our use of three different cohort construction strategies (each with a different sample size and amount of information needed) makes the possibility of failure due to inadequate sample size or event rate less likely (compared to relying on a single approach). Third, conceptually, even if undesirable, emulation failures due to inadequate sample size, low event rates, or high missingness provide valuable information about our ability to use observational data to emulate target trials. For instance, it is interesting to compare the sample sizes obtained when trying to emulate the actual trials versus when trying to emulate a target trial that would have used more liberal inclusion criteria (as in the expanded population analyses). In particular, large differences in sample size (especially if they are large enough to render the trial emulation infeasible or highly imprecise) provide information about the narrowness of the actual trial's eligibility criteria and the applicability of its findings to real-world practice.

(2) Short performance period of the award: the performance period of this project is fairly short and, given the complexity of the data management and statistical analysis tasks to be performed, there is a risk of project delays. We have taken steps to mitigate such risks by budgeting sufficient programmer and analyst time. As noted in **Section 7b**, the PI of this proposal has previous experience with OptumLabs data in the context of a project whose aims were similar to those of the current proposal. They also have a track record of working together and are committed to the success of this project. To monitor progress the entire study team will meet weekly to discuss progress, solve problems, and set short term (~week) and medium-term goals (~month). Dahabreh has a very close working relationship with Saeed (who will be implementing methods outside the OptumLabs sandbox) and Robertson (who will be working in the sandbox). Furthermore, one of our analysts (Robertson) has also had hands on experience on the OptumLabs environment. And we will enlist the effort of a programmer highly experience in administrative data analyses from the many housed in the Department of Health Services Policy and Practice at Brown University. Furthermore, we believe that our team's understanding of and ability to implement the relevant causal and statistical methods is a unique strength that will allow us to make rapid progress on the analysis of the data. That said, as a last resort, to ensure we can meet project deadlines, we might omit a small number of the proposed analytical methods (e.g., use only estimating equations instead of both estimating equations and TMLE approaches to obtain doubly robust estimators, since asymptotically these two approaches are equivalent for the causal models we describe in this proposal). We do not think this is a likely contingency.

11. Project Team Responsibilities

As partners, OptumLabs and your research project team are responsible for various aspects of research projects. By completing this document, you acknowledge and accept that your institution is responsible for the following:

1. Collaborating with OptumLabs on a SOW
2. Assuring your team follows all contracts and policies, including confidentiality obligations
3. Finalizing the research plan
4. Ensuring project feasibility
5. Management of budget, timeline, deliverables, and risks
6. Completing research project analysis as agreed upon in SOW
7. Status reporting
8. Summary of study results
9. Publications, including OptumLabs participation in reviews (see **Section 17**)
10. Leading Translation to practice and/or commercialization
11. Project termination

For more information about roles and responsibilities, please consult your institution's Master Agreement with OptumLabs.

12. Publication of Research Findings Related to an SOW with OptumLabs

OptumLabs and your institution have agreed to certain rights and obligations regarding publishing the results of research conducted at OptumLabs. Unless otherwise agreed in writing between OptumLabs and your institution, you may publish Summary Results subject to the following:

- You must send any proposed document for publication or presentation to OptumLabs at least 30 days before submission or presentation. Publications include, without limitation, the examples in **Section 5** of this document.
- OptumLabs will review the document for various items, including without limitation:
    - Confidential Information, including certain intellectual property and all individual level data, which must be removed before submission or presentation
    - Appropriate acknowledgement and description of the OptumLabs data asset and OptumLabs branding
    - Authorship recognition consistent with the International Committee of Medical Journal Editors (ICMJE) guidelines and recommendations and identification of you as a Visiting Fellow (or similar) at OptumLabs. (E.g., "Pat Smith, MD, PhD, Your-Institution, and Visiting Fellow at OptumLabs")
- In some cases, OptumLabs may require you to delay submission or presentation to allow protection of intellectual property rights.

13. Conflict of Interest

Submitters of this Research Application attest that all researchers listed do not have a conflict of interests, or an appearance of a conflict of interests, in the design of this research or the content of subsequent publications, that would be created by any funding from any commercial organization, or organization representing commercial interests, including funding for (a) other projects or programs with which any researcher is involved, or (b) activities carried out by any researcher under another affiliation or association (e.g., separate consulting business).

14. Complete list of references

1. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *American journal of epidemiology.* 2003;158(9):915-920.
2. Johnson ES, Bartman BA, Briesacher BA, et al. The incident user design in comparative effectiveness research. *Pharmacoepidemiology and drug safety.* 2013;22(1):1-6.
3. Hernan MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology.* 2008;19(6):766-779.
4. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol.* 2016;183(8):758-764.
5. Wilson J, Bock A. The benefit of using both claims data and electronic medical record data in health care analysis. *Optum Insight.* 2012:1-4.
6. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology.* 2004:615-625.
7. Patel MR, Mahaffey KW, Garg J, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med.* 2011;365(10):883-891.
8. Wang SV, Franklin JM, Glynn RJ, Schneeweiss S, Eddings W, Gagne JJ. Prediction of rates of thromboembolic and major bleeding outcomes with dabigatran or warfarin among patients with atrial fibrillation: new initiator cohort study. *BMJ.* 2016;353:i2607.
9. Nauck M, Frid A, Hermansen K, et al. Efficacy and safety comparison of liraglutide, glimepiride, and placebo, all in combination with metformin, in type 2 diabetes: the LEAD (liraglutide effect and action in diabetes)-2 study. *Diabetes Care.* 2009;32(1):84-90.
10. Chen N, Lutsey PL, MacLehose RF, et al. Association of Oral Anticoagulant Type With Risk of Dementia Among Patients With Nonvalvular Atrial Fibrillation. *J Am Heart Assoc.* 2018;7(21):e009561.
11. Noseworthy PA, Yao X, Abraham NS, Sangaralingham LR, McBane RD, Shah ND. Direct Comparison of Dabigatran, Rivaroxaban, and Apixaban for Effectiveness and Safety in Nonvalvular Atrial Fibrillation. *Chest.* 2016;150(6):1302-1312.
12. Shrestha S, Baser O, Kwong WJ. Effect of Renal Function on Dosing of Non-Vitamin K Antagonist Direct Oral Anticoagulants Among Patients With Nonvalvular Atrial Fibrillation. *Ann Pharmacother.* 2018;52(2):147-153.
13. Yao X, Abraham NS, Sangaralingham LR, et al. Effectiveness and Safety of Dabigatran, Rivaroxaban, and Apixaban Versus Warfarin in Nonvalvular Atrial Fibrillation. *J Am Heart Assoc.* 2016;5(6).
14. McHorney CA, Peterson ED, Ashton V, et al. Modeling the impact of real-world adherence to once-daily (QD) versus twice-daily (BID) non-vitamin K antagonist oral

anticoagulants on stroke and major bleeding events among non-valvular atrial fibrillation patients. *Curr Med Res Opin.* 2018:1-8.

15. Deshpande CG, Kogut S, Willey C. Real-World Health Care Costs Based on Medication Adherence and Risk of Stroke and Bleeding in Patients Treated with Novel Anticoagulant Therapy. *J Manag Care Spec Pharm.* 2018;24(5):430-439.
16. Wittbrodt ET, Eudicone JM, Farahbakhshian S, McAdam-Marx C. Comparison of low-dose liraglutide use versus other GLP-1 receptor agonists in patients without type 2 diabetes. *Am J Manag Care.* 2018;24(8 Suppl):S156-S164.
17. Dahabreh IJ, Robertson SE, Tchetgen EJT, Stuart EA, Hernan MA. Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics.* 2018.
18. Dahabreh IJ, Robertson SE, Stuart EA, Hernan MA. Transporting inferences from a randomized trial to a new target population. *arXiv preprint arXiv:180500550.* 2018.
19. Dahabreh IJ, Sheldrick RC, Paulus JK, et al. Do observational studies using propensity score methods agree with randomized trials? A systematic comparison of studies on acute coronary syndromes. *Eur Heart J.* 2012;33(15):1893-1901.
20. Kitsios GD, Dahabreh IJ, Callahan S, Paulus JK, Campagna AC, Dargin JM. Can We Trust Observational Studies Using Propensity Scores in the Critical Care Literature? A Systematic Comparison With Randomized Clinical Trials. *Crit Care Med.* 2015;43(9):1870-1879.
21. Dahabreh IJ, Kent DM. Can the learning health care system be educated with observational data? *JAMA.* 2014;312(2):129-130.
22. Gershman B, Guo DP, Dahabreh IJ. Using observational data for personalized medicine when clinical trial evidence is limited. *Fertil Steril.* 2018;109(6):946-951.
23. Danaei G, Garcia Rodriguez LA, Cantero OF, Logan RW, Hernan MA. Electronic medical records can be used to emulate target trials of sustained treatment strategies. *J Clin Epidemiol.* 2018;96:12-22.
24. Caniglia EC, Zash R, Jacobson DL, et al. Emulating a target trial of antiretroviral therapy regimens started before conception and risk of adverse birth outcomes. *AIDS.* 2018;32(1):113-120.
25. Hernan MA. Invited commentary: hypothetical interventions to define causal effects--afterthought or prerequisite? *Am J Epidemiol.* 2005;162(7):618-620; discussion 621-612.
26. Hernan MA, Sauer BC, Hernandez-Diaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol.* 2016;79:70-75.
27. Cain LE, Saag MS, Petersen M, et al. Using observational data to emulate a randomized trial of dynamic treatment-switching strategies: an application to antiretroviral therapy. *Int J Epidemiol.* 2016;45(6):2038-2049.
28. Garcia-Albeniz X, Hsu J, Hernan MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. *Eur J Epidemiol.* 2017;32(6):495-500.
29. Haavelmo T. The probability approach in econometrics. *Econometrica: Journal of the Econometric Society.* 1944:iii-115.
30. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology.* 2006;17(4):360-372.
31. Swanson SA, Hernan MA. The challenging interpretation of instrumental variable estimates under monotonicity. *Int J Epidemiol.* 2018;47(4):1289-1297.

32. Hernan MA, Robins JM. *Causal inference.* CRC Boca Raton, FL:; 2010.
33. Robins JM, Hernán MA. Estimation of the causal effects of time-varying exposures. *Longitudinal data analysis*: Chapman and Hall/CRC; 2008:547-593.
34. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology.* 2000;11(5):550-560.
35. Hernan MA, Robins JM. Per-Protocol Analyses of Pragmatic Trials. *N Engl J Med.* 2017;377(14):1391-1398.
36. Young JG, Hernan MA, Picciotto S, Robins JM. Relation between three classes of structural models for the effect of a time-varying exposure on survival. *Lifetime Data Anal.* 2010;16(1):71-84.
37. Young JG, Tchetgen EJT, Hernán MA. The choice to define competing risk events as censoring events and implications for causal inference. *arXiv preprint arXiv:180606136.* 2018.
38. Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiology.* 2001:313-320.
39. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics.* 2005;61(4):962-973.
40. Gruber S, Van Der Laan MJ. Targeted maximum likelihood estimation: A gentle introduction. 2009.
41. Gruber S, van der Laan MJ. tmle: An R package for targeted maximum likelihood estimation. 2011.
42. Chernozhukov V, Chetverikov D, Demirer M, et al. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal.* 2018;21(1):C1-C68.
43. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol.* 2008;168(6):656-664.
44. Abadie A, Imbens GW. Large sample properties of matching estimators for average treatment effects. *econometrica.* 2006;74(1):235-267.
45. Bickel PJ, Sakov A. On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica.* 2008:967-985.
46. Otsu T, Rai Y. Bootstrap inference of matching estimators for average treatment effects. *Journal of the American Statistical Association.* 2017;112(520):1720-1732.
47. Abadie A, Imbens GW. On the failure of the bootstrap for matching estimators. *Econometrica.* 2008;76(6):1537-1557.
48. Little RJ, Rubin DB. *Statistical analysis with missing data.* Vol 333: John Wiley & Sons; 2014.
49. Schomaker M, Heumann C. Bootstrap inference when using multiple imputation. *Statistics in medicine.* 2018;37(14):2252-2266.