The YODA Project
Research Proposal Review

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

The Yale University Open Data Access (YODA) Project
Yale University Center for Outcomes Research and Evaluation (CORE)
1 Church Street, Suite 200, New Haven, CT 06510

The YODA Project
Research Proposal Review - Final
(Protocol #: 2017-2511)

Reviewers:

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

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

Comments:

The applicants have answered our responses satisfactorily.

The YODA Project
Research Proposal Review

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

The Yale University Open Data Access (YODA) Project
Yale University Center for Outcomes Research and Evaluation (CORE)
1 Church Street, Suite 200, New Haven, CT 06510

The YODA Project
Research Proposal Review - Revisions Requested
(Protocol #: 2017-2511)

Reviewers:

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

Review Questions:	**Decision:**
1. Is the scientific purpose of the research proposal clearly described?	No
2. Will request create or materially enhance generalizable scientific and/or medical knowledge to inform science and public health?	Yes
3. Can the proposed research be reasonably addressed using the requested data?	Yes, or it's highly likely
4. Recommendation for this data request:	Not Approve

Comments:

The proposal was vague in its methodological intent and would be strengthened with minor revisions to be sure pre-specified plans are clear.

I appreciate the authors' intent with this innovative proposal but found some of the language to be a bit too vague. I think the proposal could be clearer to other scientists, and thus allow better understanding of the objectives and methods, with revision. For instance, the specific aims of the project are quite broadly stated. Could the investigators be more specific with the explicit intent of these analyses?

The authors' proposed statistical analysis could be better defined:
"This can be performed either in one step using a regression model with suitable interaction terms between the X's and the treatment arm Z, or separately for each treatment arm, thus being non-parametric on the form of these interactions. There is no need for the model to be correctly specified to determine an ITR. However, model misspecification may influence the properties of the resulting ITR at two levels. First, a parametric model may put too many constraints on the relationship between predictors and treatment effect, and therefore miss regions of the parameter space where one treatment in superior to the other, or conversely, wrongly identify regions where a treatment would seem beneficial, thus decreasing the benefit of the ITR overall. Second, the model for E(Y(1)|X) and E(Y(0)|X)—and therefore the ITR—and the benefit of this ITR are usually estimated using the same set of data, which leads to over-fitting and over-optimism."
The authors should consider writing the statistical analysis section to mirror the explicit research objective statement and to be clear in which is the primary analysis and how it will be performed, which is the secondary analysis and how it will be performed.

The authors explain that they will write two articles: one for a statistical audience and a second for a clinical audience. I was confused by this. Will the statistical article report a method that will be developed using the CANTATA-SU data? Or will it report a method developed prior to beginning work with this data? I'd imagine that the clinical article will be a report of the application of the method to this data. Please clarify.

Principal Investigator

First Name: Raphaël
Last Name: Porcher
Degree: PhD
Primary Affiliation: Université Paris Descartes
E-mail: raphael.porcher.pro@gmail.com
Phone number: +33 1 4234 7814
Address: Team METHODS, CRESS, Hotel-Dieu
1 place du Parvis Notre-Dame
City: Paris
State or Province: -
Zip or Postal Code: 75004
Country: France
SCOPUS ID: 55682557800

General Information

Key Personnel (in addition to PI):

First Name: Raphaël
Last name: Porcher
Degree: PhD
Primary Affiliation: Université Paris Descartes
SCOPUS ID: 55682557800

First Name: Félix
Last name: Balazard
Degree: MS
Primary Affiliation: Université Pierre et Marie Curie

Are external grants or funds being used to support this research?: External grants or funds are being used to support this research.
Project Funding Source: Felix Balazard is supported by a PhD grant from French ministry of Higher education and research (ministère de l'enseignement supérieur et de la recherche)
How did you learn about the YODA Project?: Colleague

Conflict of Interest

http://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors-rp-signed.pdf
http://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors-fb.docx

Certification

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

1. NCT00968812 - A Randomized, Double-Blind, 3-Arm Parallel-Group, 2-Year (104-Week), Multicenter Study to Evaluate the Efficacy, Safety, and Tolerability of JNJ-28431754 Compared With Glimepiride in the Treatment of Subjects With Type 2 Diabetes Mellitus

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

Research Proposal

Project Title

Policy-aware evaluation of personalized treatment strategies

Narrative Summary:

In therapeutic evaluation, the treatment with the highest response rate (or average outcome) is usually considered as superior to the others. This would however be true only if the responders to the “inferior” treatment would all respond to the “superior” one. When sets of responders do not overlap, an optimal treatment strategy could lead to a much higher response rate in the overall population
In this project, we aim at developing statistical methods to both identify individualized treatment rules targeting the responders to each treatment, and evaluate the population benefit of using such rules.

Scientific Abstract:

Background: Personalized or precision medicine aims at giving the “right treatment to the right patient”. It is one of the most promising areas of medical research, but its development is hindered by methodological limitation of studies used. Determining individualized treatment rules (ITR) is an active research field of biostatistics relying on recent statistical methods (machine learning, classification of high-dimensional data) that raises many statistical and computational challenges. For instance measures of the benefit of an ITR as compared to a ‘one size fits all’ treatment strategy where all patients receive the treatment performing best on average have been proposed. Testing whether personalization provides overall benefit after estimating an ITR however remains an open problem.
Objective: We aim at developing a statistical test for the benefit of personalization, as well as innovative approaches to identify ITRs, and to apply these methods to real data.
Study design: Retrospective analysis of randomized controlled trial.
Participants: Diabetes patients included in the trial and receiving at least one dose of study drug.
Main Outcome measure: Change in HbA1c from baseline to week 52.
Statistical analysis: The estimation of ITR will rely on modeling the outcome using treatment arm and covariates using random forests. The estimating and test of the benefit of the ITR will account for the uncertainty in the ITR estimation, and provide proper confidence intervals as well as control of the type I error rate.

Brief Project Background and Statement of Project Significance:

The objective of personalized or precision medicine is to give “the right patient the right drug at the right moment”. Precision medicine therefore implies determining which treatment is the best for a given patient, based on his/her characteristics, instead of favoring the one with better outcome on average in the whole population (one size fits all). Indeed, the current practice in medicine is to favor the treatment with the highest response rate (response being intended in the general sense of any favorable outcome). This would be a reasonable rule only if the responders to the “inferior” treatment would all respond to the “superior” one. However, when sets of responders do not overlap, an individualized treatment strategy could lead to a much higher response rate in the overall population. For instance if the usual treatment strategy only has a 20% response rate and 40% of patients respond to the new treatment, the response rate in the overall population could range from 40% (if all responders to the usual treatment respond to the new one) to 60% (if none of them respond to the new treatment).
Developing methods to identify patients more likely to respond to a treatment than to another one has recently become a very active research topic in biostatistics. Once a model predicting whether a patient would be more likely to respond to a given treatment or to its comparator has been obtained, for example using data from a randomized controlled trial (RCT), it is straightforward to derive an individualized treatment rule (ITR) where patients would receive the treatment under which their predicted response is higher. It has been shown that such a strategy would maximize the expectation of the outcome over the population.
Measures of the performance of an ITR using a biomarker as compared to the “one size fits all” strategy have also been developed, such as the improvement in population average outcome under the ITR, for instance. The

specification of the model to estimate the performance of the ITR is however important, especially when several biomarkers are considered together. The classical approach relies on generalized linear regression with markers by treatment interactions, but the effect of model misspecification can be critical, especially around the decision boundary. This would lead to biased predicted performance of ITR. Even using flexible approaches such as machine learning, there remains a non-negligible risk to recommend the incorrect treatment for patients with close predicted response under each treatment. In addition, testing whether personalization provides overall benefit remains an open problem.

Statement of project significance
Our work is important for two main reasons. First, it is crucial that responders to each treatment compared would be correctly identified to develop individualized treatment strategies that will improve the outcome of patients. In that respect, there is a need for cutting-edge statistical methods. Second, it is also important that the benefit of such individualized strategies would not be overstated and overestimated because there is a risk of false decision at potentially high costs.

Specific Aims of the Project:

The aims of this project are:
(1) To develop statistical tests for the benefit of personalization.
(2) To develop innovating approaches to determine individualized treatment strategies using data from randomized controlled trials.
(3) To extend the proposed approaches to the situation of more than two treatments.
To our knowledge, all methods for modeling individualized treatment strategies in the literature have considered a two-treatments setting. Nevertheless, the real-life situation is different, and for most diseases, there are more than two therapeutic alternatives.
(4) To illustrate the potential gain of the approach we propose using real data from randomized controlled trials. The request for data to the Yoda platform primarily serves this aim. As a corollary, this will allow ultimately to determine individualized treatment strategies for diabetic patients with an associated measure of potential population benefit.

What is the purpose of the analysis being proposed? Please select all that apply.
New research question to examine treatment effectiveness on secondary endpoints and/or within subgroup populationsPreliminary research to be used as part of a grant proposalOther

Research Methods

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

Selection of the trials:
The methods we develop need RCTs with relatively large sample size, as well as relevant patients characteristics. Based on these considerations, we have selected the study NCT00968812 (CANTATA-SU trial) as a possibly good candidate for our methodology, since it has a large sample size, the experimental and control treatments have different mechanisms of action, which is likely better suited for finding variables associated with a differential treatment effect, and the effect of the experimental treatment as compared to the comparator is not overwhelming, thus allowing for a more refined strategy.

Selection of the patients:
All patients included in the selected trial and receiving at least one dose of study drug (modified intent to treat analysis as reported in the study primary reports) will be considered.

Main Outcome Measure and how it will be categorized/defined for your study:

Our main outcome will be the same as the primary study outcomes, i.e. change in HbA1c from baseline to week 52.

To illustrate the potential of the method for a binary outcome, which has been more frequent in statistical articles on the issue of individualized treatment strategies, we will add a binary key secondary outcome, which will be the proportion of patients achieving HbA1C <7.0% (53 mmol/mol).

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

The primary independent variable is the treatment arm allocated. Since the CANTATA-SU comprises three treatment arms, we will use as primary comparison the comparison of canagliflozin 300 mg + metformin versus glimepiride + metformin. In a second stage, we will perform a similar analysis for canagliflozin 100 mg + metformin versus glimepiride + metformin.

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

The other variables of interest are the patient characteristics that will be used to construct a model for individual treatment. In order to capture at best the heterogeneity in treatment effect, as many baseline (pre-randomization) variables as possible should be considered in our models. We list here a minimal set of variables that could be used:
Sex
Age
Race
Glycated haemoglobin A1c (HbA1c)
Fasting plasma glucose (FPG)
Bodyweight
Body-mass index
Duration of type 2 diabetes
Whether patient entered antihyperglycaemic drug adjustment period
Smoking history
Other diseases or comorbidities (whenever available)
Systolic blood pressure
Diastolic blood pressure
Pulse rate
Triglycerides
LDL cholesterol
HDL cholesterol
Non-HDL cholesterol
Insulin
Alanine aminotransferase
Aspartate aminotransferase
Alkaline phosphatase
Bilirubin
Blood urea nitrogen
Gamma-glutamyltransferase
Urate
Hemoglobin
Urine albumin/creatinine
Total fat mass
Total lean mass
Subcutaneous adipose tissue
Visceral adipose tissue

Statistical Analysis Plan:

The analysis will consider a counterfactual or potential outcomes framework, where we posit that for each patient, there exists two potential outcomes, namely Y(1) and Y(0), representing the outcome that the patient would experience should s/he receive the studied treatment (indexed by 1) or its comparator (indexed by 0), respectively. This allows defining a counterfactual individual treatment effect as D = Y(1) – Y(0). In practice, D cannot be observed, except under very specific trial designs such as n-of-1 trials. The question of precision medicine is thus rather to estimate the expected value of D given a set of covariates X. Assuming that higher values of Y represent a more favorable outcome, it has been shown that the optimal individualized treatment rule given X—or optimal treatment regime—corresponds to give treatment 1 to patients with $E(D|X) > 0$ and treatment 0 to patients with $E(D|X) < 0$. There is equipoise for those with $E(D|X)=0$, so that the decision to favor one of the treatment should be based on other considerations, such as being conservative and favoring the ‘older’ treatment for instance. Several measures have been proposed to quantify the benefit of personalization of a treatment, i.e. the benefit of

using an individualized treatment rue as compared to a non-personalized treatment strategy where all patients receive the same treatment whatever their covariates X (one size fits all). We will rely on the improvement in population average outcome under the ITR, i.e. the difference in expectations of the outcome Y in the population under the ITR strategy vs a one size fits all strategy. In situations where the new treatment is superior to its comparator on average, this ITR implies giving the new treatment to all patients except if E(D|X) ? 0. This measure accounts both for the improvement in average outcome among those who benefit from the ITR (i.e. those for whom E(D|X) ? 0), and their proportion in the total population.
To find an ITR from observed data, we have to rely on an estimation of E(D|X). Since D is not observed, methods rely on estimating E(Y(1)|X) and E(Y(0)|X) using regression modeling. This can be performed either in one step using a regression model with suitable interaction terms between the X's and the treatment arm Z, or separately for each treatment arm, thus being non-parametric on the form of these interactions. There is no need for the model to be correctly specified to determine an ITR. However, model misspecification may influence the properties of the resulting ITR at two levels. First, a parametric model may put to much constraints on the relationship between predictors and treatment effect, and therefore miss regions of the parameter space where one treatment in superior to the other, or conversely, wrongly identify regions where a treatment would seem beneficial, thus decreasing the benefit of the ITR overall. Second, the model for E(Y(1)|X) and E(Y(0)|X)—and therefore the ITR—and the benefit of this ITR are usually estimated using the same set of data, which leads to over-fitting and over-optimism. To derive corrected point and interval estimates and statistical tests overcoming these issues, Janes et al (2014) have proposed a double bootstrap procedure for instance. Nonetheless, this remains computer-intensive, and performs rather poorly under the null, which has led to a two-stage approach where interactions are tested first, before estimating the benefit of the ITR.
In our work, we rely on modeling the outcome using treatment arm and covariates using random forests, in order to allow more flexibility in the outcome model, and derive a proper procedure to for point and confidence interval estimation, as well as statistical testing, accounting for the uncertainty in the individualized treatment effect estimation.
In this project, missing outcome and predictor values will be handled through multiple imputation by chained equations.

Project Timeline:

We are currently working on the methodological developments and performing simulation studies to investigate the properties of our procedure in realistic settings. Analyzing the trial should be straightforward once they are in an analysis-ready format. Depending on the format of data provided, however, this could imply additional data management tasks. We however plan to have the analyses ready in 6 to 8 months.

Dissemination Plan:

Our primary purpose is to illustrate how the methods we develop perform in real settings. To this aim, we plan to draft a first article for a statistical journal such as JASA or Biometrics, where the data would serve as illustration only.
Then we plan to draft also a clinical article for a medical audience (in a journal such as BMJ, PloS Medicine, BMC Medicine, or a specialty journal such as Diabetes or Diabetes Care), where the results of the study would be presented for non-statisticians, expecting a clinical impact of our project.

Bibliography:

Cai T, Tian L, Wong PH, Wei LJ. Analysis of randomized comparative clinical trial data for personalized treatment selections. Biostatistics 2011; 12:270–282.
Foster JC, Taylor JM, Ruberg SJ. Subgroup identification from randomized clinical trial data. Stat Med 2011; 30:2867–2880.
Huang EJ, Fang EX, Hanley DF, Rosenblum M. Inequality in treatment benefits: can we determine if a new treatment benefits the many or the few? Biostatistics 2017; 18(2):308–324.
Huang Y, Fong Y. Identifying optimal biomarker combinations for treatment selection via a robust kernel method. Biometrics 2014; 70(4):891–901.
Huang Y, Laber EB, Janes H. Characterizing expected benefits of biomarkers in treatment selection. Biostatistics. 2015;16(2):383–99.
Janes H, Brown MD, Huang Y, Pepe MS. An approach to evaluating and comparing biomarkers for patient treatment selection. Int J Biostat 2014; 10(1):99–121.
Janes H, Pepe MS, McShane LM et al. The fundamental difficulty with evaluating the accuracy of biomarkers for

guiding treatment. J Natl Cancer Inst 2015; 107(8):djv157.
Kang C, Janes H, Huang Y. Combining biomarkers to optimize patient treatment recommendations. Biometrics 2014; 70(3):695–707.
Li J, Zhao L, Tian L, Cai T, Claggett B, Callegaro A, Dizier B, Spiessens B, Ulloa-Montoya F, Wei LJ. A predictive enrichment procedure to identify potential responders to a new therapy for randomized, comparative controlled clinical studies. Biometrics. 2016; 72(3):877–887.
Lipkovich I, Dmitrienko A, Denne J, Enas G. Subgroup identification based on differential effect search - a recursive partitioning method for establishing response to treatment in patient subpopulations. Stat Med 2011; 30:2601–2621.
Porcher R, Jacot J, Biau D. Identifying treatment responders using counterfactual modeling and potential outcomes. Presented at EpiClin 2016, manuscript submitted.
Qian M, Murphy SA. Performance guarantees for individualized treatment rules. Ann Stat 2011; 39(2):1180–1210.
Shalit U, Johansson F, Sontag D. Estimating individual treatment effect: generalization bounds and algorithms. arXiv:1606.03976; 2016.
Shen J, Wang L, Taylor JMG. Estimation of the optimal regime in treatment of prostate cancer recurrence from observational data using flexible weighting models. Biometrics 2017;73(2):635–645.
Shen J, Wang L, Daignault S, Spratt DE, Morgan TM, Taylor JMG. Estimating the optimal personalized treatment strategy based on selected variables to prolong survival via random survival forest with weighted bootstrap. J Biopharm Stat 2017 (Ahead of print).
Su X, Tsai CL, Wang H, Nickerson DM, Li B. Subgroup analysis via recursive partitioning. J Mach Learn Res 2009; 10:141–158.
Zhang B, Tsiatis AA, Laber EB, Davidian M. A robust method for estimating optimal treatment regimes. Biometrics 2012; 68:1010–1018.
Zhao L, Tian L, Cai T, Claggett B, Wei LJ. Effectively selecting a target population for a future comparative study. J Am Stat Assoc 2013;108:527–539.
Zhao YQ, Zeng D, Laber EB et al. Doubly robust learning for estimating individualized treatment with censored data. Biometrika 2015; 102(1):151–168.
Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. J Am Stat Assoc. 2017 (ahead of print).