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 #: 2020-4390)

Reviewers:

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

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

Comments:

No additional comments.

The YODA Project
Research Proposal Review

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

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

The YODA Project
Research Proposal Review - Revisions Requested
(Protocol #: 2020-4390)

Reviewers:

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

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

Comments:

The authors aim to apply a novel statistical approach that they developed on two clinical trials. Can the authors outline which outcomes, from the two trials that were identified, will be selected? Are the main outcomes from both trials appropriate for the analyses (i.e, are they continuous)?

I understand the rationale here but I do wonder whether the results are actionable: if we find heterogeous effects how can we ultimately apply that clinically? Aren't we better off by looking at clinically relevant subgroups or other clinical predictors of response (biomarkers, demographics, etc)?

Principal Investigator

First Name: Mohamed Aziz
Last Name: Mezlini
Degree: PhD
Primary Affiliation: Massachusetts General hospital. Harvard medical school
E-mail: azizmezlini@gmail.com
Phone number: 6477196678
Address: 65 Landsdowne. Cambridge. Partners Health lab
65 Landsdowne
City: Cambridge
State or Province: MA
Zip or Postal Code: 02139
Country: United States

General Information

Key Personnel (in addition to PI):
First Name: Mohamed Aziz
Last name: Mezlini
Degree: PhD
Primary Affiliation: Harvard Medical school. Massachusetts General hospital

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

Conflict of Interest

https://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors_2019_11.pdf

Certification

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

1. NCT00216476 - RISSCH3001 - CONSTATRE: Risperdal® Consta® Trial of Relapse Prevention and Effectiveness
2. NCT00679627 - GALALZ3005 - A Randomized, Double-Blind, Placebo-controlled Trial of Long-term (2-year) Treatment of Galantamine in Mild to Moderately-Severe Alzheimer's Disease

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

Research Proposal

Project Title

A novel statistical approach to detect the efficacy of drugs that have heterogeneous treatment effects

Narrative Summary:

We developed a novel statistical test that is particularly empowered to detect drug efficacy in scenarios where the drug works well on a small proportion of individuals. We would like to apply our test to clinical trials to showcase its usefulness.

Scientific Abstract:

Background: The statistical tests currently used in clinical trials to prove drug efficacy are looking for a difference in mean between intervention groups. Therefore, these methods can be underpowered in a setting with heterogeneous treatment effects. We developed a novel statistical test that is better suited for these situations.
Objective: We aim to apply our test on several clinical trials to showcase its usefulness. The goal is to prove that we have better power at detecting efficacy compared to traditional approaches.
Study design: For each clinical trial studied, we will use our test to assess whether there is a significant difference between cases and controls in the main outcome measures. Our test is empowered to detect heterogeneous treatment effects.
Participants: All participants in the clinical trials requested.
Outcome measure: Power to detect efficacy (with small enough statistical p-value) as a function of the sample size in comparison with traditional approaches.
Statistical analysis: We will use our novel statistical test.
(https://www.biorxiv.org/content/10.1101/2020.03.23.002972v2 publication under review)

Brief Project Background and Statement of Project Significance:

There is often a lot of variability among patients with the same diseases, this includes variability in disease stage, variability in the underlying disease causes and pathways, and variability in drug metabolism profiles. For all these reasons, a given drug might work in some patients and show no effect on many others. This is called a heterogeneous treatment effect.
Clinical trials are often conducted with the aim of proving the drug's efficacy. However, the statistical methods frequently used in these trials are looking for an average effect and are not specifically designed to detect a heterogeneous treatment effect scenario. Therefore these methods can sometimes be underpowered to detect drug efficacy.
We designed a novel statistical test for this scenario and showed that it has superior power compared to widely used statistical tests: https://www.biorxiv.org/content/10.1101/2020.03.23.002972v2
Better power means potentially the discovery of more beneficial drugs and/or the need for fewer participants and lower costs for future trials.
We aim to show the usefulness of our test in real data of clinical trials.
The novel statistical test can be useful for any setting with heterogeneous effects. In the future, it can be applied in other clinical trials as well as other
research fields such a biomarker discovery.

Specific Aims of the Project:

We aim to show the performance of our statistical test on real clinical trials. We will use the data to demonstrate superior power to detect efficacy under heterogeneous treatment effects compared to traditional tests previously used in these trials such as ANCOVA, Wilcoxon, LMM, etc.
The power is the ability to detect the association between the intervention and the outcome with a significant p-value, given sample size.
The null hypothesis: There is no difference between the cases (individuals receiving the drug) and controls (placebo or standard care).
The alternative hypothesis: The distribution of the outcome is different between cases and controls, with a significant proportion of cases benefiting from the drug.

What is the purpose of the analysis being proposed? Please select all that apply.
Develop or refine statistical methods

Research Methods

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

No exclusion criteria

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

We will test the main outcomes of each study for association with the intervention using our test.
In each clinical trial, we will use the same main outcome as the original analysis (found in clinicaltrials.gov and the associated publications) provided the outcome is continuous. In studies having both discrete and continuous main outcomes, we will only test the continuous outcome.

The outcome of our analysis is an assessment of the power to detect the drug efficacy (using p-vaues) and a comparison with the traditional approaches that were used in the trial.

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

In each clinical trial analyzed, the main predictor will be the Intervention (drug, placebo).
We are looking at a case-control setting. If there are more than two groups (for example multiple drug dosages), we will restrict the analysis to two groups at once, doing the same pairwise comparisons as the original analysis.

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

We will correct for the exact same covariates (such as gender, age and others) as was done in the original analysis of each clinical trial (details were found in both clinicaltrials.gov and the associated clinical trial publication).

Statistical Analysis Plan:

We went through the list of trials on clinicaltrial.gov and curated trials that fit the following criteria: A large enough sample size, a continuous main outcome, and the potential for heterogeneous treatment effects (by looking at the summary statistics).
The reason for this selection is that we want trials where our test is applicable and where we can hope to have enough power to detect drug efficacy. We requested the clinical trials at the intersection of our list and the list of trials available on Vivli.
We will apply our novel statistical test on the main outcomes of each trial we are granted access to.
Here is a reference to our test with more details: https://www.biorxiv.org/content/10.1101/2020.03.23.002972v2
In each study, we will correct for the same covariates and interactions (same model) as was done in the original analysis of the trial. We will handle missing data (imputation or exclusion) in the same way and using the same methods as the original analysis whenever possible.
We do not plan on doing an analysis of subgroups based on covariates.
We are following a case-control design. In scenarios with more than two groups (such as multiple drug dosages), we will consider the same groupings and we will do the same pairwise comparisons as the original analysis.
Our goal is to detect efficacy as a significant statistical difference in distribution between cases and controls. Our main result will be a p-value per outcome, corrected for multiple hypothesis testing.
We will run our analysis using Vivli's platform. All analyses will be performed by Aziz M. Mezlini.
Software Used:
RStudio
Project Timeline:

Once we are granted access to the data (anticipated in August 2020), we will run our analysis and get results within 3 months (anticipated November 2020).
We will prepare a publication and a report back to Yoda Project by February (2021).
We anticipate that the whole project will take less than 12 months from the data access is granted.

Dissemination Plan:

We intend to submit the manuscript to a wide audience journal with interest in Statistics, Computational biology,

and clinical trials.
Depending on how good the results are, our first choice is currently Nature Methods but that is still open to change.

Bibliography:

https://www.biorxiv.org/content/10.1101/2020.03.23.002972v2