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 #: 2018-2801)

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:

I note the responses which help to clear up some of the points raised. Overall, I think it is reasonable to approve the request.

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 #: 2018-2801)

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?	Unsure, further clarification from requestor is needed
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:

1. The stated hypothesis is that "new safety signal detection approach have increased power of identify safety signals at both SOC level and AE level comparing to currently available safety signal detection approaches ". However, the analytic plan as written does not test this hypothesis.

2. The main analysis is to be conducted using a "newly developed score test approach", yet this approach is not described.

3. The primary study outcome is stated as "grades of all experienced adverse events". Please clarify.

4. I fully support the requestors' aims but there are some obscurities in the language, e.g. "Re-dig the value of failed drugs and/or wiser use of approved drugs". The choice of the abiraterone studies also introduces a lot of complexity as it will hard to distinguish the possible adverse effects of the drug from the effects of disease progression.

Principal Investigator

First Name: Xianming
Last Name: Tan
Degree: PhD
Primary Affiliation: University North Carolina at Chapel hill
E-mail: tan.xianming.888@gmail.com
Phone number: 9199628610
Address: 450 West Drive

City: Chapel hill
State or Province: North Carolina
Zip or Postal Code: 27599
Country: United States

General Information

Key Personnel (in addition to PI):
First Name: Xianming
Last name: Tan
Degree: PhD
Primary Affiliation: UNC Chapel Hill

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?: Internet Search

Conflict of Interest

http://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors_xmt.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
Associated Trial(s):

1. NCT00638690 - A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study of Abiraterone Acetate (CB7630) Plus Prednisone in Patients With Metastatic Castration-Resistant Prostate Cancer Who Have Failed Docetaxel-Based Chemotherapy
2. NCT00887198 - A Phase 3, Randomized, Double-blind, Placebo-Controlled Study of Abiraterone Acetate (CB7630) Plus Prednisone in Asymptomatic or Mildly Symptomatic Patients With Metastatic Castration-Resistant Prostate Cancer

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

Research Proposal

Project Title

Development and applications of new safety analysis methods for randomized clinical trials

Narrative Summary:

Drug safety is an important issue. Important safety analysis questions include: (a) safety signal detection: to identify safety issues worthy further exploration, (b) subgroup analysis: to identify subjects vulnerable to certain safety issues, (c) benefit-risk analysis.
We have developed a series of new analytical approaches to tackle well-known challenges including (1) high dimensional, (2) rare occurrence, and (3) weak signal . In this project, we will (1) apply these approaches to detect safety signals of different drugs based on drug clinical trial data, and
(2) develop and test subgroup analyses approaches to identify subgroups with different benefit-risk profiles.

Scientific Abstract:

Background: Drug safety has always been a major medical concern, and adverse drug events are associated with cost of multi billions annually to the health delivery system. Yet, current practice of safety analyses in clinical trials are at most descriptive, featuring a long list, running possibly hundreds of pages, of observed adverse events (AEs) grouped by treatment arms, and relying on ocular investigation.An urgent need exists for the development of advanced approaches, which could be formally incorporated in the design and analysis of clinical trials, including cancer clinical trials, for safety evaluation.
Objective: Our primary objective is to apply several powerful safety signal detection approaches to existing drug trial data. Secondary objectives include developing and applying subgroup analysis approaches to identify subgroups with different benefit-risk profiles.
Study Design: It is a secondary data analysis project.
Participants: Our analysis will include all subjects available from these data sets we have applied for access. Analyses will be conducted separately for each trial, or on combined trial data if same drug was tested in different trials.
Main Outcome Measure: The primary endpoint will be safety outcome (incidence of AEs) for the first objective (safety signal detection). Efficacy outcomes will be added for the secondary.
Statistical Analysis: The primary endpoint will be analyzed using a hierarchical testing approach. We will develop new approaches for subgroup analyses.

Brief Project Background and Statement of Project Significance:

Safety issues of a drug can include occurrence of specific AEs as well as clinically significant lab tests or other safety-related measurements. The safety of marketed drugs has become a public concern since the Elixir Sulfanilamide disaster in late 1930s which led to the Federal Food, Drug and Cosmetic Act (FD&C Act) (1938), and then Kefauver-Harris Drug Amendments (1962) as a response to the Thalidomide tragedy. The past decades have seen numerous regulatory guidance documents issued to enhance drug safety evaluation by both national and international organizations like International Conference on Harmonisation (ICH)1, Council for International Organizations of Medical Sciences (CIOMS)2, Food and Drug Administration (FDA)3, European Commission (EC)4.

Safety assessment is thus critically important in drug development. Astonishingly contrary to this importance is a lack of well-accepted, and statistically sound safety assessment approach in clinical trials, due to several statistical challenges including multiplicity, rare events, and weak signal. Current popular practice of safety analysis only involves descriptive statistics, ending with a lengthy, but arguably useless report of frequencies of all observed safety issues during the study period of a clinical trial. Failure to adjust for multiplicity increases the risk of false positives, while an overly stringent adjustment can result in false negatives, a real troublesome issue from a product-consumer perspective. The fact that most trials are designed based on efficacy endpoints and may have limited power to detect safety signals makes the situation even worse.

On the other hand, around 50% new drug failed at Phase III, posing great financial burdens to drug developers. In addition, even for approved drug, the existence of population heterogeneity largely implies sub-optimal use of drugs. Re-dig the value of failed drugs and/or wiser use of approved drugs could benefit from appropriate subgroup

analyses based on benefit-risk properties of these drugs.

The proposed research is significant for its potential to greatly enhance current safety evaluation practice in drug development. Current frequentist approaches in this area have overly focused on multiplicity using different false discovery rate control methods5, but leaves the issues of rare events and weak signal largely untouched. The project will show that a new frequentist approach framework, which implicitly accounts for correlated AEs, together with classical false discovery rate control methods, could well tackle the safety evaluation challenges. In addition, more advanced subgroups analyses approach with emphasis on latent population heterogeneity could lead to re-activation of 'dead' drugs and optimal use of approved drugs. This could eventually lead a well-established safety evaluation and subgroup analyses framework for in drug trials, and, through preventing approval of unsafe drug, minimizing preventable cost, caused by drug safety issue, to public, government, and pharmatheutical industry.

Specific Aims of the Project:

We hypothesize that our new safety signal detection approach have increased power of identify safety signals at both SOC level and AE level comparing to currently available safety signal detection approaches like double false discovery rate (FDR).
Primary specific aim / objective:
1. To identify system organ class (SOCs) and individual AEs which show statistically significant difference between treatment arms, with appropriate control of type I error.
Secondary specific aims / objectives:
1. To develop new subgroup analysis approaches to account for both safety and efficacy outcomes.
2. To apply the subgroup analysis approaches to identify subgroups subgroups who are vulnerable to certain AEs, and subgroups who might benefit from a study drug without excessive toxicity.
3. To generalize the above approaches to analyze data pooled from different trials on the same drug.

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 populationsNew research question to examine treatment safetyPreliminary research to be used as part of a grant proposalParticipant-level data meta-analysisParticipant-level data meta-analysis using only data from YODA Project

Research Methods

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

We will employ a post-hoc analysis of the COU-AA-301/302 data to compare the joint distribution of safety outcomes between treatment arms to identify safety signals. Our primary endpoint is the multivariate safety outcomes observed in a trial, and our secondary outcome will add efficacy outcomes.
All patients treated on the COU-AA-301/302 trial will be eligible for this analysis. Requested Data:
Demographic information
-Age
-Race
-Gleason Score
-Date of Diagnosis
-Prior anti-cancer therapies
-Prior prostatectomy and/or radiation therapy
-Investigations (PSA, Hgb, Cr, AlkPhos, LDH)
-Pain score
-Performance Status
-Cohort
-#cycles
-Mode of progression
-Best PSA response
-CTCAE (grades of all experienced adverse events)
-Date of initiation
-Date of PSA progression (PSA progression-free survival)
-Date of Radiographic PFS (Radiographic progression-free survival)
-Date of death (overall survival)

-CT and NM Bone scan description of metastatic disease at baselin (# bone, nodal, visceral)
-CT and NM Bone scan description of metastatic disease at progression (# bone, nodal, visceral)

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

The primary endpoint is the grades of all experienced adverse events. Thus, the main outcome measure will be the grades of all experienced adverse events. Efficacy outcomes like time of disease progression, overall survival will be considered as secondary outcomes for subgroup analyses.

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

The primary independent variable will be assigned study treatment (abiraterone versus placebo). It is assigned per study protocol.

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

Age at study entry (years)
Race (Am. Indian, Asian, Black , Native Hawaiian or Other Pacific Islander, White)
Gleason Score (sum)
Date of Diagnosis
Prior anti-cancer therapies (number of prior hormonal therapies, prior ketoconazole, prior chemotherapies)
Prior prostatectomy and/or radiation therapy (Y/N for each)
Investigations (PSA, Hgb, Cr, AlkPhos, LDH)
Pain score / presence of pain (binary Y/N)
Performance Status (ECOG)
Cohort (Abiraterone or Placebo)
cycles administered / duration of exposure to agent (in # of cycles)
Mode of progression (clinical, radiographic, toxicity)
Best PSA response (% reduction)
Date of Abiraterone or Prednisone initiation
Date of PSA progression (PSA progression-free survival)
Date of Radiographic PFS (Radiographic progression-free survival)
Number of bone mets at baseline
Number of nodal mets at baseline
Number of visceral mets (liver, lung, other) at baseline
Number of bone mets at progression – and if progression in bone
Number of nodal mets at progression – and if progression in node
Number of visceral mets (liver, lung, other) at progression – and if progression in viscera

Statistical Analysis Plan:

Disease characteristics will be compared using descriptive statistics and t-tests. Difference in safety profile between treatment arms per SOC will be tested using a newly developed score-test approach. Difference in incidence of individual AEs between treatment arms will be conducted using Fisher’s exact test, adjusted for multiplicity using a hierarchical testing approach.

Subgroup analysis will be based on new approaches to be developed. Specifically, we pan to first apply tree-based approaches (e.g., decision tree, random forest) to build classification/regression trees for safety and efficacy outcomes, so as to identify potential subgroups with outstanding safety/efficacy profiles. We also plan to extend current tree-based approaches to handle potential issues like latent heterogeneity in study population, e.g., there could exist unobserved (latent) factors contributing to the definition of safety/efficacy subgroups.

Project Timeline:

Project start date: 7/2018
Analysis completion date: 7/2019
Date manuscript drafted/submitted: 7/2020
Results reported 7/2020

Dissemination Plan:

Anticipate presentation of data at JSM 2019 or JSM 2020, with manuscript publication in Statistics in Medicine, Clincial Trials, or similar journal.

Bibliography:

1. http://www.ich.org/products/guidelines/safety/article/safety-guidelines....
2. http://www.cioms.ch/publications/blurbs/management_of_safety_information...
3. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformati...
4. http://ec.europa.eu/health/files/eudralex/vol-10/2011_c172_01/2011_c172_...
5. Mehrotra DV1, Adewale AJ. Flagging clinical adverse experiences: reducing false discoveries without materially compromising power for detecting true signals. Stat Med. 2012 Aug 15;31(18):1918-30. doi: 10.1002/sim.5310.
6. Sun, J., Zheng, Y., and Hsu, L. (2013) A Unified Mixed-Effects Model for Rare-Variant Association in Sequencing Studies. Genet Epidemiol. 2013 Mar 9. doi: 10.1002/gepi.21717
7. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011;89:82–93.