The YODA Project
Research Proposal Review

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

The YODA Project
Research Proposal Review - Final
(Protocol #: 2019-3936)

Reviewers:

- ☒ Nihar Desai
- ☒ Joshua Wallach
- ☐ 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:

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 #: 2019-3936)

Reviewers:

- ☒ Nihar Desai
- ☒ Joshua Wallach
- ☐ 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:

This is an interesting proposal to apply an item response theory framework for adjusting patient centered outcome-based scores ascertained from clinical trials for patient drop-out that occurs at random. While it is clear that this statistical approach is typically used for quality of life measures, the authors do not specify which quality of life measure they intend to use for each of the 6 trials requested. It would improve the proposal if the authors were to provide this information.

It is unclear whether the requested data will actually be used for the proposed analyses: “Hence, I have requested a variety of trials of this type in hopes of finding one that is appropriate”. It appears the researchers have selected cancer or mental disorder trials with missingness not a random and a fairly large sample size, but it would be helpful if the researchers clarified their rationale for the selected trials.

It would be useful if the researchers are able to specifically reference the outcomes of interest in the available CSR summary files. These can be used to determine primary and secondary outcomes. The authors note “Because collected variables reflecting PCO often are not part of primary or even secondary endpoints, it’s not necessary clear which datasets will be appropriate for analyses”. If the outcomes are not listed in the CSRs, how will access to IPD change that? Do these data sets contain the kind of patient outcome data that would make this statistical approach viable?

This revised data request addresses a number of the questions raised after the first review of the proposal. However, it is still unclear whether the requested trials will fulfill the eligibility criteria. The investigators note that the trials must have missing not at random as part of the design, with a substantial propotion of drop-out. It seems as if the investigators could have reviewed some of the trial specific information regarding drop-out (e.g. ClinicalTrials.gov or the corresponding publications) to determine potential eligibility. That being said, if the investigators are able to identify an eligible trial, this work has the potential address a gap in the statistical literature.

Principal Investigator

First Name: Charles
Last Name: Iaconangelo
Degree: PhD
Primary Affiliation: Pharmerit, LP
E-mail: ciaconangelo@pharmerit.com
Phone number: 6096022789
Address:

City: Brooklyn
State or Province: NY
Zip or Postal Code: 11201
Country: USA

General Information

Key Personnel (in addition to PI):
First Name: Daniel
Last name: Serrano
Degree: PhD
Primary Affiliation: Pharmerit, LP

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

Conflict of Interest

https://yoda.yale.edu/system/files/ds_yoda_project_coi_form_for_data_requestors_2019.pdf
https://yoda.yale.edu/system/files/cji_yoda_project_coi_form_for_data_requestors_2019.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. NCT00589914 - R092670PSY3006 - A Randomized, Double-Blind, Parallel-Group, Comparative Study of Flexible Doses of Paliperidone Palmitate and Flexible Doses of Risperidone Long-Acting Intramuscular Injection in Subjects With Schizophrenia
2. NCT00638690 - COU-AA-301 - 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
3. 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
4. NCT00887198 - COU-AA-302 - 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
5. NCT02076009 - 54767414MMY3003 - Phase 3 Study Comparing Daratumumab, Lenalidomide, and

Dexamethasone (DRd) vs Lenalidomide and Dexamethasone (Rd) in Subjects With Relapsed or Refractory Multiple Myeloma

6. NCT02136134 - 54767414MMY3004 - Phase 3 Study Comparing Daratumumab, Bortezomib and Dexamethasone (DVd) vs Bortezomib and Dexamethasone (Vd) in Subjects With Relapsed or Refractory Multiple Myeloma

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

Research Proposal

Project Title

Unbiased Treatment Efficacy Detection Methods with Patient Centered Outcomes

Narrative Summary:

Patient Centered Outcomes (PCO) are often incorporated in trials via patient/clinician/caregiver reported questionnaires. In the survival analysis context, such as cancer trials, patients drop-out of the study for reasons related to disease severity. This type of drop-out leads to incorrect conclusions about patient quality of life. This research develops statistical approaches that are needed that appropriately adjust for this missingness and yield correct inferences. More robust development of these methods will help detect which interventions have a positive impact on patient quality of life. This, in turn, will help guide patient centered drug development.

Scientific Abstract:

Background: Patient Centered Outcomes (PCO) are often incorporated in trials via patient/clinician/caregiver reported questionnaires. In the survival analysis context, the data is typically considered to be missing not at random (MNAR), which requires the use of statistical approaches that properly adjust for this type of missing data.
Objective: This research presents an item response theory (IRT) framework for adjusting PCO-based scores for MNAR drop-out. The model allows the IRT scores and the drop-out mechanism to be modeled simultaneously. This restores conditional independence and corrects for bias in the estimates of treatment efficacy that occur under MNAR drop-out.
Study Design: A simulation study was designed, run, and analyzed to illustrate the improved estimates of treatment efficacy using this approach. An empirical example using clinical trial data demonstrates the utility of the procedure.
Participants: Patients from a clinical trial with a survival analysis design will be included in the analysis.
Main Outcome Measures: In the simulation study, the bias and root mean squared error (RMSE) of the estimated separation in the treatment arms will be computed. In the empirical data example, the treatment arm separation will be estimated with the proposed procedure and without adjustment. Differences in the estimated separation and resulting inferences will be reported.
Statistical Analysis: This will compare the estimates of treatment efficacy from the proposed longitudinal IRT model with the estimates of treatment efficacy from a standard IRT model.

Brief Project Background and Statement of Project Significance:

Statistical models in the psychometrics literature have been slow to penetrate the field of Patient Centered Outcomes (PCO). Latent variable models, of which IRT is a sub-category, offer advantages in modeling data with measurement error.1 Although there has been interest in applying the IRT framework to PCO data, there have been barriers to widespread adoption. Specifically, there is a lack of easily implemented procedures for dealing with MNAR data. Little methodological work has been done to address this issue since the initial research on the topic.2 Until these methods are developed and validated, IRT-based estimates of treatment efficacy using PCO will be biased under MNAR data. This work addresses a gap in the statistical literature by developing a psychometric model that can accommodate MNAR data. Adapting modern psychometric methods for the field of PCO is a crucial step in addressing the lack of sensitivity in many PCO measures.3

Specific Aims of the Project:

There are 4 specific aims of the project: 1. Theoretical statistical exposition of methods under development. 2. Simulation studies of proposed methods. 3. Evaluation of proposed methods using randomized clinical trial data. 4. Publication of proposed methods, findings, and software from simulation and empirical studies.

Aim 1: See attached manuscript for outline of the theoretical exposition (equations are in LaTeX and won't render here).

Aim 2: Simulation studies demonstrate the improved estimates of treatment efficacy compared to other methods. This work hypothesizes that the proposed method will return estimates with smaller bias and RMSE than standard approaches. This will lead to greater power to detect treatment arm separation.

Aim 3: After showing that the procedure can recover the true parameters in a simulation study, the utility of the proposed method will be demonstrated using data from a clinical trial. The clinical trial should be a survival analysis design, which makes it highly likely to feature MNAR drop-out mechanisms. Applying the proposed method should then yield different estimates than an unadjusted IRT model.

Aim 4: This work will be published in a statistics journal. Additionally, all software code written will be made available online.

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:

The data source will be one that contains a robust set of PCO in order to allow illustration of the statistical method. Because collected variables reflecting PCO often are not part of primary or even secondary endpoints, it's not necessarily clear which dataset will be appropriate for analysis. One important criterion is that the data be MNAR – this typically occurs in a survival analysis context. Cancer trials and mental disorder trials typically feature this type of drop-out mechanism. Hence, I have requested a variety of trials of this type in hopes of finding one that is appropriate.

If more than one clinical trial is appropriate, then a determination will be made if those trials can be combined into one analysis set. For example, if both studies of Daratumumab meet these criteria, these two studies can be combined and analyzed together. This is important because the IRT framework typically requires larger sample sizes. If the trials cannot be combined, then the most appropriate trial will be selected. This will likely entail selecting the trial with the largest sample size.

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

Note here that the proposed method is the reverse of the typical survival analysis. For example, in oncology trials, the typical approach compares time to drop-out across treatment arms, with an adjustment for patient health-related quality of life. The work here does the opposite: the longitudinal IRT model compares quality of life across treatment arms, with an adjustment for drop-out. This assumes that the drop-out is related to the patient quality of life – that is, it assumes the data is MNAR due to the survival analysis design. The ultimate purpose of the proposed method is to compare patient health-related quality of life across the treatment arms, without the bias that occurs from ignoring the MNAR drop-out mechanism.

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

More broadly, IRT models use questionnaire items to evaluate the construct of interest. Any and all items related to that construct of interest can be incorporated in a typical IRT model. The proposed approach goes one step further and also incorporates patient drop-out. That is, the patient data is re-coded to reflect the timepoint at which they dropped out. To summarize, item response on the PCO questionnaire and patient drop-out time will be utilized in the model.

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

Standard practice is to include basic demographic information, such as age and sex, in the statistical model. This ensures that potential confounds have been controlled for, which allows for valid interpretation of the model output.

Statistical Analysis Plan:

The proposed use of this dataset is to illustrate a newly developed statistical method. This illustration will be part of a manuscript that provides a technical exposition of the methods, as well as a simulation study to show the performance of the method. A minimal amount of descriptive statistics will be computed, with the main focus being a comparison of the estimated treatment effect using the proposed method versus standard methods. To accommodate this analysis, descriptives such as the proportion of drop-out at each time point, as well as the average score at each timepoint (stratified across treatment arms) will be computed. This will help to show how the method makes adjustments for missing data and how that impacts the model-based estimates of treatment efficacy.
Software Used:
R
Project Timeline:

A draft manuscript has been attached as a file below. A technical exposition has been sketched out, and initial simulation evidence already compiled. The application to the clinical trial data will help guide the simulation study conditions. The project will start once the data have been transferred.
Timeline:
Data management: 1 month
Application to clinical trial, creating tables/figures, writing up results: 3 months
Full simulation study, creating tables/figures, writing up results: 3 months
Completing writing, revisions: 3 months
Submission to journal: 10 months from data transfer

Note: typically journal reviewers will ask for additional work to be done. Given the 6-18 month review times for popular statistical journals, an extension will almost certainly be necessary.

Dissemination Plan:

Please see attached file for a draft of a manuscript. This is being prepared for the journal Statistics in Medicine. The goal is to highlight psychometric methods that can and should be utilized in medical applications. Just as importantly, software code will be disseminated. The code will be included as an Appendix to the manuscript, available online. The same code will also be posted to a github account, making it easily searchable. Please note that this software code will NOT include any sensitive information regarding the requested dataset.

Bibliography:

1. de Ayala RJ. The Theory and Practice of Item Response Theory (Methodology in the Social Sciences). New York: The Guilford Press; 2009.
2. Douglas JA. Item response models for longitudinal quality of life data in clinical trials. Stat Med. Nov 15 1999;18(21):2917-2931.
3. Lawrence Gould A, Boye ME, Crowther MJ, et al. Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. Statistics in Medicine. 2015;34(14):2181-2195.

Supplementary Material:

https://yoda.yale.edu/sites/default/files/irt_mnar_draft_06202019.pdf