The YODA Project
Research Proposal Review - Revisions Requested
(Protocol #: 2019-3840)

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?	Unsure, further clarification from requestor is needed
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:

1. This is an interesting proposal to use data from 3 trials to predict new atrial fibrillation among patients receiving ibrutinib. My concern is that the data the investigators have requested are not prepared as part of a de-identified patient data set. The Data Generator has confirmed that while ECG datasets are available, scans cannot be properly de-identified and consequently cannot be made available. Please update your proposal accordingly.

2. My request for clarification relates to the validation part of the proposal. It seems the data from 3 trials are being used to develop a predictive algorithm to that identifies patients at higher risk for new atrial fibrillation when receiving ibrutinib. The investigators then explain that they will test this algorithm on a historical control data source. However, this data source is never described. A few details would be useful. What data source is this?

3. Please clarify if both ECG and echocardiographic measures will be used as part of algorithm development (in addition to clinical and laboratory parameters), as most of the text seems focused only on the ECG measures.

4. My main concern about this proposal centers on study power. If the Afib rate is 10%, how many new onset Afib will they have? Parenthetically, using SAE/AE terms to find Afib will be challenging. It is not adjudicated (so you don't know if the patient really had Afib) and not systematically collected. On top of this, with limited "N" how can you really have so many covariates?