

**The YODA Project**  
**Research Proposal Review - Final**  
**(Protocol #: 2022-4949 )**

**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:**

1. In the proposal, the first paragraph of background is talking about leveraging clinical trial data with synthetic controls to calculate relative effectiveness and risk of AEs on a patient-level. Is it referring to the use of patient-level data to assess relative effectiveness and risk of AEs or assessing those at patient-level?
2. The proposal mentions splitting the data into training set and test set for the modeling efforts. There does not seem to be details on how to split, except that the proposal mentions about use the different percentage. Also, depending on how the models are constructed, they may need a training set, a cross validation set, as well as a test set. Without knowing how the models are being constructed, it is hard to tell whether the approach is reasonable.
3. The objective is saying to validate the synthetic control method and apply the method to quantify the treatment effect. Please clarify if the validation of the model is referring to the validation of the model and not the model itself.
4. In the specific aims, only validation is mentioned and in the statistical analysis plan, it also seems to be the case that validation is what this proposal is doing. If that's the case, the objective in the earlier part needs to match these.
5. In the statistical analysis plan, it is calling continuous variables as continuous values and binary variables as binary values. These, again, are mixing the concept of variables and values. Please clarify.