**Reviewer 1 Comment**

I have only one concern about the setting of the paper. It seems for the historical datasets, we need datasets from both RCTs and an external control.  
How about there is a suitable historical RCT without an external control?  
How can we construct an external control for such a RCT?  
I hope the authors can discuss this point in the paper.

**Response**

We’ve added this paragraph into the Discussion section to discuss this point.

*Our approach also requires generating an external control arm for each reference study. Because we are adopting propensity scoring to account for differences in measured covariates between the study and external control populations then it is likely that the same external control data source could be used for each study with different propensity score weights calculated to reflect differences in the populations between the reference studies. This is how the Flatiron Health database was used in the example in section 4. There may be situations where the overlap in population between the external control data source and one or more of the reference studies is too small to permit a meaningful estimation of treatment effect in which case, if other more relevant external data sources for that study are unavailable, then the best course of action may be to remove these studies from the reference set. However this is only likely to happen for studies performed in a substantially different population to the rest of the reference studies.*