Major Compulsory Revisions

1. The authors described in page 10 that matching is conducted by examining whether the CI for a unit in the treatment group overlaps with that for a unit in the control group. If one to one matching is applied, the method will take the unit in the control group which has the closest distance from the unit in the treatment group. Can the authors explain what they mean by “closest distance”? Is there an algorithm used to determine “closest” distance—such as Mahalanobis distance? For example, The CI of a treatment unit, [CI:2,5] is overlapped with those of two control units, [CI for control A: 1.5, 5] and [CI for control B:2,5.5]. Which control unit has the closest distance from this treatment unit?

2. The authors compared their proposed approach with nearest neighbor matching and Caliper matching. The nearest neighbor matching is known in the PSM literature as a problematic matching method (Caliendo and Kopeinig Journal of Economic Surveys 2008), and provided the worst balance in the covariates between groups (DÕAGOSTINO Stats in Medicine 1998 and Baser Value in health 2006). By contrast, Mahalanobis matching with caliber provided the best balance. It will be of interests to compare the matching balance the new method proposed by the authors with Mahalanobis matching with caliper. Given that matching balance between their proposed approach is very similar with Caliper method, this new comparison might help to decide whether their new approach is superior to current approaches.

3. Interval matching is known as stratification matching in the PSM literature. This method also deals with the range of variation of the propensity score. Can the author explain further the main difference between their proposed approach and the stratification matching? It will be also interesting to see the comparison in the matching balance between their proposed method and stratification matching.

4. Current PSM approaches often use common support to check the overlap of the distribution between two groups. I wonder whether common support is needed in their approach.

5. The authors stated that the proposed approach can be easily implemented in one to many matching. However, in the situation of one to many matching, analysis used in the matched sample should incorporate weights that reflect the ratio/probabilities of treatment and control units sharing similar PSM point estimate. Can the authors explain in such situation how weighting can be
incorporated in their proposed approach?
6. It is very impressed to consider 74 covariates in the PSM application. Can the authors describe how they deal with missing data in these covariates?

Minor Revisions:
1. The authors stated in page 3 that “estimate a propensity score for each unit using a logistic regression of treatment conditions on covariates;”. However, logistic regression is not the only approach to estimate propensity scores. Probit, semi-parametric, non-parametric regression and decision tree are alternatives of estimating propensity scores.

2. The authors stated in the abstract that “interval matching reduced slightly more than did caliper matching”. However, Table 2 and Figure 2 show that both interval matching provides similar good balance with caliper matching.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests