# Minutes of the Population-Level Estimation Workgroup

April 6, 2016, Eastern Hemisphere meeting

Dukyong Yoon, Sung Jae Jung, Rae Woong Park, 10 students in South Korea, Jenna Reps, Peter Rijnbeek, Ian Wong, Nicole Pratt, Martijn Schuemie

(I’m sorry, I didn’t get the names of all students. Their majors included Internal medicine, computer engineering, and biomedical informatics.)

Martijn reopened the workgroup discussion by presenting his slides. In summary, Martijn argued that the results produced by observational research so far as reported in scientific literature are highly suspect. The primary reasons are (1) study bias, (2) publication bias, and (3) p-hacking. The objective of this workgroup is to clean up this mess, or more formally:

*Develop scientific methods for observational research leading to population level estimates that are accurate, reliable, and reproducible, and facilitate the use of these methods by the community.*

Martijn listed many topics of interest for the workgroup, including:

* Best practices for estimation studies
* How to present and interpret results from estimation studies
* Should we not do single studies anymore?
* Should humans make analysis choices, or do we let the data decide?
* Overview of the current methods library
* Evaluation of methods
* Training on methods
* Funding and collaboration opportunities
* Whatever comes up for discussion

The definition of best practices is something that is expected to grow over time. Martijn has created a framework document that we expect to fill in through discussions in the workgroup:

<http://www.ohdsi.org/web/wiki/doku.php?id=development:best_practices_estimation>

Ian remarked that he often does the opposite of the p-hacking example in Martijn’s slides: when he encounters p< .05 he’ll try different adjustments to see how robust that is. Even though we agree it is probably just as bad as the other type of p-hacking, we can’t always avoid it altogether.

Dukyong hypothesized that if we were extract estimates not only from the abstract but also the full text we might see more estimates with p > .05. This is something we could check.

Nicole agreed with the suggestion that we should always (try to) include a pre-exposure risk window when performing an SCCS. However, she’s not sure she would like to recommend a one-size-fits-all length for this window. Instead, she’d like to do research into how to determine the optimal length for specific situations.

Nicole and Martijn went on a bit of a tangent and discussed the fact that even though the SCCS design relies on several assumptions holding true, there are no formal tests for these assumptions (unlike other designs, where we for example can test for covariate balance or whether the proportionality assumption holds in a Cox regression). Nicole expressed keen interest in developing such tests.