**Minutes of the Population-Level Estimation Workgroup**

November 16, 2016

Present: Seng Chan You, Ajou Students, Danny Kim, YG Kim, Sukhoon Lee, Martijn Schuemie, Yuriy Khoma

## Part 1

Seng Chan You presents on the conversion of the Korean nationwide health insurance (NHI) database to the CDM. One of the goals is to establish a drug even monitoring system like sentinel, EU-ADR, MIHARI (Japan), and CNODES. NHI in Korea started in 1977, from 1985 covers almost every person in Korea. There are two DBs in Korea: NHIC and HIRA database. NHIC doesn’t have hospitalization, HIRA doesn’t have demographics. Annual health checkups are captured.

Started definining the ETL using a 1M person (2%) sample to OMOP CDM V5. The whole DB is estimated to take 15TB. There are no mappings for measurements codes.

A pilot study on the full database is defined for two activities: (1) cohort extraction, and (2) can we find known ADRs? Cohort extraction was tested on anti-HBC (Entecavir) using ATLAS. The know ADRs are:

* ADR: ACE inhibitors and dry cough. Using CohortMethod package. PS matching on age and sex only due to restricted computer power at government.
* ADR: AC inhibitor and angioedema. Same process as before. Wide confidence interval
* ADR: Olmesartan and enteropathy

## Part 2

Seng Chan You presents on an ongoing study on the 1M person sample, comparing combinations of treatment in hypertension. The guidelines allow for combinations of drugs. The outcome is ‘Major adverse cardio/cerebral event’. It was challenging to create cohorts for combinations of drugs. Furthermore, 40% of drugs not mapped to RxNorm, but to ATC instead. Added ATC to concept\_ancestor table (e.g. cetapril).

Post matching there were still major differences between the cohorts. Had to include Charlson + DCMI scores to make comparable. After this, the cohorts were comparable, even on lab values not included in the propensity score! Martijn suggests always including as many covariates as possible in the propensity model. The regularization will deal with there being too many variables (unless the number of subjects is < 2,500).