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

November 2, 2016

Present: Rae Park, Kenneth Man, Sanghyung Jin, Ajou students, Soo Yeon Cho, Sukhoon Lee, YG Kim, Hojun Park, Nicole Pratt

Today we had two presentations. The first, on **Distributed Degression**, was presented by Sungjae Jung, a Master’s student in Biomedical Informatics at Ajou University. The second, on **Biosignal Database Contruction for the Development of Disease Prediction Models**, was presented by Dukyong Yoon, assistant professor in the same group as Sungjae.

# Distributed regression

We have data distributed over ten hospitals (or other databases in OHDSI) which we cannot pool, but we would like to fit a single regression model. One possible solution is Grid Binary Logistic Regression (GLORE), which works by integrating decomposable partial elements. The Newton method for optimizing the likelihood needs only the first and second derivative of likelihood, so we can communicate these to a central cite, and send updates to beta in return. GLORE was adapted in SCANNER project, and Sungjae’s idea is to also use it OHDSI. Perhaps we can have a ‘Super WebAPI’ that coordinates amongst local WebAPIs to run a distributed regression.

For Sungjae’s graduation project, Sungjae is focusing on the idea that GLORE can be applied in a single database as well, when that database is too large too fit into memory. The only real restriction would be disk access. For this purpose, Sungjae will implements the GLORE algorithm in an R package.

Martijn asks whether GLORE can also be applied to Cox models, since then it could be used to compute ‘meta-analysis’ estimates for comparative effectiveness questions. Sungjae mentions the WebDISCO algorithm, where this has been done before, so yes, it should be possible.

**Biosignal Database Contruction for the Development of Disease Prediction Models**

The Bioinformatics group received a $500,000 grant over 3 years from the Korea Health Industry Development Institute to build prediction models for sepsis using bio-signals. Sepsis is a serious condition, with 17% mortality, 34% developing a septic shock.

The prediction algorithm uses deep learning. Typical EMR data is too sparse, hence the need for bio-signal data such as ECG, repiratory data, SpO2, blood pressure, temperature, CO2, CO, EEG, etc. Ajou University is now collecting and storing all this information, a huge amount of data (data storage capacity is currently 200TB). Ajou has worked on connecting Nihon-Kohden and Philips monitoring devices to the database. GE devices use a closed protocol, so are excluded from the analysis. Currently, data from 40 devices is captured on a continuous basis. Data capture started on September 1, 2016. In the near future, other devices such as the ventilators and continous lab data will be added. 1 sep 2016 starting gathering real data.

Rae stresses the need for integration of this information in the CDM, and also mentions that Ajou is working on a real-time CDM, where data is added to the CDM as it is collected in the hospital.

Next meeting is November 16.