**Exploring the Required Number of Items for Kampo Pattern Prediction Models Using Machine Learning**

**Ayako Maeda-Minami1,2**, Tetsuhiro Yoshino2, Yuko Horiba2, Kenji Watanabe2, Koichi Fukunaga2.

Faculty of Pharmaceutical Sciences, Tokyo University of Science1, Katushika-ku, Tokyo, Japan;

Faculty of Medicine, Keio University2, Shinjuku-ku, Tokyo, Japan.

**Background and aims.** Kampo is a traditional medicine in Japan. Kampo diagnosis is part of the traditional East Asian medicine classification included in ICD-11, and Kampo treatments are selected based on the individual patient's pattern. We have previously reported on a prediction model using a machine learning model for Kampo pattern diagnosis, which required a large number of questionnaire items. Pharmacists can sell over-the-counter Kampo medicines, however, not all pharmacists are well educated in Kampo and can determine the symptoms. To enable pharmacists to more easily select appropriate over-the-counter Kampo medicines, a predictive model that can predict Kampo patterns using fewer items is needed. The objective of this study was to explore a predictive model for Kampo pattern diagnosis that requires fewer questionnaire items.

**Methods.** We collected data from patients’ first visit to one of the six Kampo clinics in Japan from January 2012 to February 2015. All participants were surveyed using a 153-item questionnaire, including 5 vital signs and 148 subjective symptoms. We constructed prediction models of deficiency-excess and cold-heat pattern diagnosis using a random forest algorithm from training data and extracted the most important items. To examine whether pattern diagnosis can be identified using a smaller number of items, we extracted the top 3-30 most important items based on importance rankings and constructed machine learning models using only these selected items.

**Results.** The number of eligible patients was 750. In the deficiency-excess pattern prediction model, the most important items were BMI, systolic blood pressure, diastolic blood pressure, age, and easily fatigued, in that order. In the cold heat pattern prediction model, the most important items were cold hypersensitivity in legs, cold intolerance, cold hypersensitivity in hands, general cold hypersensitivity, and heat intolerance, in that order. The discriminant ratio of the prediction model using all 153 questionnaire items was 80.8% for deficiency-excess pattern and 74.9% for cold-heat pattern. When using only the top five items, the discriminant ratios were 78.0% and 73.6%, respectively.



**Figure 1.** Discriminant ratio of pattern prediction models constructed using the top 3 to 30 most important questionnaire items and all items.

**Conclusion/Discussion.** We confirmed an early saturation of the discriminant ratio of a pattern diagnostic prediction model. We plan to continue our research to facilitate the selection of appropriate herbal medicines.

**Acknowledgement:** This work was supported by a Grant-in-Aid for Research on Propulsion Study of Clinical Research from the Ministry of Health, Labor and Welfare in building the questionnaire and data collection.