Assessment of Oxidative Stress and Metabolic Risks With Machine Learning Algorithms in Ethnically Diverse Population: Cross-Sectional Data in Korea and the Netherlands

Yunsoo Kim,[1] Tim Van Den Broek,[2] Jildau Bouwman,[2] Oran Kwon,[3] and Edith J M Feskens[4]

[1]Ewha Womans University, Department of Nutritional Science and Food Management; [2]Netherlands Organization for Applied Scientific Research (TNO); [3]Ewha Womans University; and [4]Wageningen University & Research

Objectives: The underlying mechanisms for the development of chronic lifestyle-related diseases result from enhanced oxidative stress, and method for quantifying in healthy adult individuals have been investigated in the previous study. However, this predictor targeted the Korean population, and it is currently unclear how to evaluate oxidative stress in diverse ethnicities. On top of that, impaired metabolism is another major process, which is closely related to cellular, organ, and organism levels interacting with oxidative stress to initiate these diseases. Therefore, this study is aimed to develop and validate a predictor to quantify oxidative stress and metabolic risks across the ethnicity.

Methods: This study was conducted using cross-sectional data of the Nutrition Questionnaires plus (NQ plus) study in the Netherlands ($n = 618$) and the Korea National Health and Nutrition Examination Survey (KNHANES) ($n = 6{,}387$). To merge these two separate population data equally, propensity score matching was applied. Based on health space statistical methodology, each axis was obtained from two different binary samples: healthy controls ($n = 324$) versus oxidative disease cases ($n = 391$) and metabolic syndrome cases ($n = 73$). The model development was carried out in 10-fold cross-validation of ridge, elastic-net, and LASSO (least absolute shrinkage and selection operator) regularized regression. Evaluating the predictive accuracy of the fine-tuned model was validated by unmatched data after propensity score matching.

Results: A total of 22 and 18 features were selected in fine-tuned models to distinguish oxidative stress and metabolic risks from healthy controls respectively. The performance of the prediction was assessed with the area under the receiver operating characteristic curve of 0.972 (CI 0.967 to 0.976) and of 0.962 (CI 0.954 to 0.97).

Conclusions: These practical composite biomarkers appear useful to stratify oxidative stress and metabolic risks in healthy adults beyond differences between populations. Further external validation is required to ensure prognostic value and general applicability in longitudinal cohorts.

Funding Sources: This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science & ICT(2012M3A9C4048761).