How to evaluate added value of new bio- and genetic markers over the conventional risk factors/markers for better prediction of patient's clinical phenotypes?

L.J. Wei
Harvard University

Recent technology advancements for obtaining bio- and genetic-markers have drastically enhanced the knowledge of certain disease processes and the potential for accurately predicting patient's clinical outcomes. Traditional statistical methods for the so-called individualized/ personalized medicine with such markers are derived under a rather strong assumption, that is, one can accurately identify the true model (at least for the large sample case), which relates the predictors to their corresponding clinical phenotype variable(s). In practice, however, it is difficult if not impossible, even to identify the class of models which contains the true one. Therefore, it is interesting and important to investigate whether the standard statistical methods for model estimation, evaluation and comparisons can be modified when the fitted model may not be correctly specified. In this talk, we discuss new procedures for predicting future observations and for evaluating and comparing prediction rules. One key feature of the proposals is that their validity does not require that assumption that the fitted models are correct. Moreover, the new proposal provides a reliability measure of the estimated prediction precision, an important component for model evaluation and checking. The new methods are illustrated with examples with continuous, binary and censored responses. We also use these examples to show how to estimate the added value from bio- and genetic markers over the routinely obtained clinical markers for predicting the clinical outcomes. Lastly even if, on an average sense, the markers are useful (or not useful), it is important to identify subgroup of patients who would benefit from the new markers. We will discuss new proposals how to locate such a subgroup.