Risk prediction models in cancer—statistical learning viewpoints

<u>熊昭</u>1

¹National Health Research Institutes

Abstract

In this talk, I will compare three statistical methods to build absolute risk models for predicting cancer occurrence using three examples. BCRAT, the well-known Gail model predicting breast cancer occurrence, is constructed using nested case-control dataset embedded in and composite hazard ratio from a large cohort and forms the framework for the development of competing risk models for predicting cancer occurrence. PLCO m2012, an absolute risk model predicting lung cancer in the next 6 years for an ever-smoker, is a logistic regression model constructed using all the ever-smokers in the control arm of the PLCO trial. TNSF-SQ, an absolute risk model constructed by our group predicting lung cancer in the next 6 years for a never-smoking female in Taiwan, is a logistic regression model constructed using the case-control components of GELAC , Taiwan biobank served as part of the controls and age-specific population incidence rates derived from *Cancer Registry*. Bias-variance tradeoff will also be discussed. Other risk models will be mentioned.

Keyword: absolute risk, bias-variance tradeoff, cancer occurrence, risk prediction, statistical learning