On criteria for evaluating risk prediction models for public health applications

Ruth Pfeiffer*, National Cancer Institute

Abstract

We propose and study novel criteria to assess the usefulness of models that predict risk of disease incidence for screening and prevention, or the usefulness of prognostic models for management following disease diagnosis. The proportion of cases followed, PCF(p), is the proportion of individuals who will develop disease who are included in the proportion p of individuals in the population at highest risk. The proportion needed to follow-up, PNF(q), is the proportion of the general population at highest risk that one needs to follow in order that a proportion q of those destined to become cases will be followed. We also propose the integrated PCF, iPCF, and iPNF, the integrated PNF, obtained by integrating PCF and PNF over a range of values of q or p. Under the assumption that the risk model is well calibrated PCF, PNF, iPCF and iPNF can be estimated based on observed risks in a population alone. When the risk models are not well calibrated, PCF, PNF, iPCF and iPNF can be estimated consistently from case control data when the outcome prevalence in the population is known, and from cohort data, with baseline covariates and observed health outcomes. The criteria are illustrated with novel models that predict incidence of endometrial and breast cancer.

Keywords: Discrimination; Discriminatory accuracy; Lorenz curve.

* Presenting author