

Association of polygenic risk scores with breast cancer

Sara Urru

University of Padova, University of Torino

Breast cancer is the most widespread neoplasia among women. The best screening method up to now is mammography, but in the last years its efficiency was questioned because of over-diagnosis. A need to define a more appropriate breast cancer risk arises to better classify women and personalize their screening exam. Recent Genome-Wide Association Studies identified some common genetic variants in the population associated with breast cancer. In particular, 77 single nucleotide polymorphisms (SNPs) were combined in a polygenic risk score (PRS). For the completion of this thesis suitable multivariable statistical techniques have been identified and then applied to study genetic and standard risk factors in a case-control study for breast cancer prevention. In this study 384 among 26600 women between 46 and 67 years old who were attending mammographic screening in Biella and Turin were considered, including 117 cases and 267 controls. Two kinds of data were collected and analysed: genetic data derived from DNA sequencing and information about personal, hormonal, reproductive and family history of the women, who were required to fill a questionnaire. Regarding genetic data the genotype of the 77 PRS-SNPs and of other 3 SNPs associated with breast cancer prognosis was obtained for every woman. The first step was the imputation of missing values, so multinomial logistic regression was applied. The second step was the calculation of the PRS for every woman. Two different scores were computed: a first one using the odds ratios found in literature (PRS-77) and a second one applying the cross-validation method on our data (PRS-80). For the answers to the questionnaire, univariate logistic regressions were employed to verify the association with breast cancer, and then multivariable logistic regressions were performed including the two PRSs separately. To build the best model according to the AIC on our data a stepwise logistic regression was performed for variable selection. ROC curves were then plotted, and AUC were calculated to assess the prediction accuracy. Both PRSs were selected by stepwise regression and improved the prediction accuracy of the model based only on standard risk factors.