Zentrum für Experimentelle Medizin
Institut für Medizinische Biometrie und Epidemiologie

Prof. Dr. Heiko Becher

Martinistraße 52
20246 Hamburg
Tel:(040) 7410-59551 (Sek)
-59550 (direkt)
Fax: (040) 7410-57790
h.becher@uke.de
www.uke.de

Seminar Epidemiologie

23. Juni 2016, 16:00
Seminarraum 213 , Gebäude N55, UKE

Building stable multivariable risk scores in GWAS consortia with partially overlapping SNP data

Anne-Sophie Stöhlker (1), Livia Maccioni (2), Aslihan Gerhold-Ay (3), Alexandra Nieters (2), Martin Schumacher (1) and Harald Binder (3)
(1) Institute for Medical Biometry and Statistics, Medical Center - University of Freiburg, Faculty of Medicine
(2) Center for Chronic Immunodeficiency, Medical Center - University of Freiburg, Faculty of Medicine
(3) Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center Johannes Gutenberg University Mainz

Anne-Sophie Stöhlker, Universität Freiburg

Regularized regression approaches, such as the lasso or componentwise boosting, can be used for building genetic risk scores from case-control single nucleotide polymorphism (SNP) data. Compared to more established univariate testing strategies, the main advantage of such multivariable regression approaches is that correlation structures are already taken into account at the stage of automatic SNP selection. Compared to univariate strategies, one apparent disadvantage is that missing SNP values cannot be handled automatically, i.e. for each SNP some value needs to be available for every individual. Yet, only partial overlap of the SNPs may exist when data from several sub-studies are to be combined, e.g. in a consortium that pools genotyping data from different platforms.

We propose a re-formulation of the lasso and of componentwise boosting that resolves such issues. This is motivated by respective data from the InterLymph consortium on risk factors for non-Hodgkin's lymphoma development, where even after imputation only partial overlap of SNPs was provided for a pre-specified set of candidate genes in several sub-studies. Specifically, we use this data to illustrate a variant of componentwise boosting that requires only pairwise covariances, which can still be determined given partial overlap. This is combined with resampling for determining stable sets of SNPs as genetic risk scores. Thus, as there is no additional computational cost, the proposed approach can be recommended for multivariable building of genetic risk scores whenever there is only partial overlap of SNP data obtained from different studies in a consortium.

Organisation: Prof. H. Becher