Inference for covariates from combined family and case control data

Ruth Pfeiffer

This work was motivated by a genetic association study for cutaneous melanoma (CM), the major cause of skin cancer mortality world wide. Major risk factors for CM include cutaneous and pigmentary characteristics, sun exposure and reactions to sun exposure. Several high-risk melanoma susceptibility genes have been identified to date, including CDKN2A and CDK4. A low-risk melanoma-susceptibility gene that has been shown to be associated with CM is the human melanocortin-1 receptor gene (MC1R), that regulates pigment formation and thus determines phenotypic risk factors. To investigate the impact of MC1R variants and sun exposure on melanoma risk in subjects from North-Eastern Italy, we used information on 649 subjects, including 183 sporadic cases and 179 healthy controls from a case-control study, and 84 familial cases and 203 unaffected relatives from 55 melanoma-prone families, defined as having at least two members with melanoma. The subjects from both studies were recruited at the same hospital and examined by the same dermatologist.

As the subjects from the case-control and the family study came from the same region and study protocol and data collection methods were identical, combining both studies seems a sensible approach to improve power to detect genetic associations for rare genetic variants. We propose two methods to jointly analyze data from the case-control and the family studies to estimate effects of the various MC1R variants and their interactions with other risk factors on melanoma risk. First, we use a random effects model to specify the joint probabilities for family members and for individuals in the case control-study. We also employ two-phase sampling ideas to estimate marginal gene effects on melanoma risk from both family and case-control data.