EXPLOITING EPIDEMIOLOGICAL FINDINGS WHEN ASSESSING BIOLOGICAL DISEASE MARKERS USING PEDIGREE DATASETS

Desmond Campbell[1], Sabine Landau[2], Joanne Knight[1] and Pak Sham[3]

[1] *Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, UK*
[2] *Department of Biostatistics and Computing, Institute of Psychiatry, King's College London, UK*
[3] *Genome Research Centre, University of Hong Kong, Hong Kong*

The identification of biological disease markers for multifactorial disorders such as depression or schizophrenia remains an important aim in mental health research. At the same time rich data sources exist for mental health disorders. Disease pedigree datasets, originally collected for linkage studies, enable the separation of genetic and environmental factors. And a wealth of information on population parameters such as effects of known risk factors, heritability or age of onset curves, has been gathered through previous epidemiological research.

We report on work in progress on developing and implementing statistical methods for maximally exploiting existing information when assessing environmental factors and biological disease markers using pedigree datasets. Such approaches are hoped to lead to power gains and should become increasingly powerful as future research reveals more about each disorder. We propose employing a two-stage approach to statistical analysis. The initial or synthesis stage converts observed pedigree data into disease liability distributions, and a further analysis stage uses liability predictions as inputs to look for associations with potential disease markers. The approach can provide separate predictions for environmental and genetic liabilities and the ability to use predictions of just those aspects that bear upon a research question of interest, may result in extra power. Finally the two-stage approach brings pragmatic benefits in terms of data reduction and synthesis, facilitation of the use of existing statistical genetics tools and improving comprehensibility.

So far we have developed and implemented a prototype disease liability generator for the synthesis stage. This liability generator uses Gibbs Sampling to generate a pedigree's joint environmental and genetic liability distribution given the pedigree's disease states and observed genetic and environmental risk factors. The underlying statistical model is a modified genetic threshold model popular in behavioural genetics. The current version of the software assumes that population parameters are known from existing research. Here we demonstrate the disease liability generator, discuss possible clinical uses of the synthesized information and outline further steps to complete our two-stage approach.