Use of R in Genetic Epidemiology Designs

Jing Hua Zhao[1*]

1. MRC Epidemiology Unit, Institute of Metabolic Science, Box 285, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0QQ, UK

*Contact author: jinghua.zhao@mrc-epid.cam.ac.uk

Keywords: Genetic Epidemiology, Study Design, Power Calculation

Statistical genetics or genetic epidemiology is essential in unravelling the genetic and environmental influences on common diseases and/or quantitative measurements in human population. The major strategies for study of these complex traits have been analyses of familial aggregation, segregation, linkage and association, with the latter two relying on the availability of genetic markers. While advance of high throughput genotyping technologies has contributed significantly to the recent success in such studies, adequate study designs remain to be critical.

With a brief introduction of genetic epidemiology and statistics, I will elaborate R implementations for several scenarios including family, case-control, case-cohort, staged designs and Mendelian randomisation studies. I will provide updated results for those as reported in an earlier paper, as well as practical use and empirical results which has clearly demonstrated the need of more analytical work, Connection will be further made to other available software, which encapsulates extensions to be made in the near future.

The presentation has been motivated from our work on design and analysis of several large epidemiological cohorts and also complementary to a pre-conference tutorial on genetic analysis of complex traits. The examples should add to the list of successful stories of the R environment in research work.

References

Armitage P, Colton T. (2005). Encyclopedia of Biostatistics, Second Edition. *Wiley.*

Zhao JH (2007). **gap**: genetic analysis package. *J Stat Soft*, 23(8), 1–18.