A General Approach for Power Calculations for the Haseman-Elston Method

Hädicke O[1], Zhang H[2], Ziegler A[1]
[1]*Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Deutschland*
[2]*Department of Epidemiology and Public Health, Yale University School of Medicine, USA*

To unravel the genetic basis of complex genetic disorders, intermediate quantitative traits are often analyzed in practice instead of the clinically relevant outcome [Ziegler, König 2006]. In family studies, these are often investigated with a sample of nuclear families, consisting in at least two offspring and their parents. Risch and Zhang [2] proposed a new approach for sample size and power calculations with the Haseman-Elston method [3] as tool for analysis. They started off with the simple Falconer model [4]

$x_{it} = \mu + g_{it} + e_{it}$,

where x_{it} is the trait value of offspring t, t = 1, 2, in family i, i = 1, …, n. μ is the general mean, g_{it} is the genetic effect attributed to the diallelic major locus. Polygenic and environmental effects are absorbed in the error term e_{it}. The genetic effect g takes on values a, d, and –a, when the offspring is homozygous for the high allele, heterozygous, and homozygous for the low alleles, respectively. The Haseman-Elston method then regresses the sib-pairs squared trait difference on the proportion τ_i of alleles shared identical by descent (IBD) and the probability z_{1i} for sib-pair i sharing one allele IBD:

$y_i = (x_{i1} - x_{i2})^2 = \alpha + \beta\ \tau_i + \gamma\ z_{1i} + \varepsilon_i$

Haseman and Elston [3] have shown that the slope coefficient β is given by [3] $\beta = -2(1-2\theta)^2\ \sigma_g^2$, where θ is the recombination fraction between the marker and the trait locus, and σ_g^2 is the genetic variance attributable to the trait locus. Formulae for σ_g^2 can be found elsewhere [see, e.g., ref. 1].

For power and sample size calculations, Risch and Zhang [2] considered an additive genetic model, i.e., $\gamma = 0$, and a completely informative genetic marker for the null hypothesis H_0: $\sigma_g^2 = 0$ against the one-sided alternative H_1: $\sigma_g^2 > 0$.

In this presentation, we firstly demonstrate in Monte-Carlo simulation studies that the power calculated by Risch and Zhang do not match their theoretical levels. We argue that the hypothesis $\sigma_g^2 = 0$ used by Risch and Zhang is inadequate. We next develop new formulae for power and sample size calculations for the Haseman-Elston method based on H_0: $\theta = ½$ versus H_0: $\theta < ½$. Finally, we illustrate the validity of our new formulae in Monte-Carlo simulation studies.

In conclusion, power and sample size calculations for the Haseman-Elston method should not be performed with the formulae developed by Risch and Zhang [2]. For this purpose, one should either utilize the approach proposed by Amos et al. [5] or our new method. For this, it is important to stress the limitations of these methods. Both are restricted to a single diallelic major trait locus. However, while Amos et al. assume absence of polygenic components and/or shared environmental effects, our approach is based on the existence of a completely informative genetic marker. We are convinced that the latter limitation is less crucial because today's genome-wide linkage studies utilizing 10,000 or more single nucleotide polymorphisms lead to almost complete marker information at any chromosomal position.

References

[1] Ziegler A, König IR. A Statistical Approach to Genetic Epidemiology. Heidelberg: Wiley-VCH; 2006.
[2] Risch N, Zhang H. Extreme Discordant Sib Pairs for mapping Quantitative Trait Loci in Humans. Science 1995; 268:1584-1589
[3] Haseman JK, Elston RC. The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 1972; 2: 3-19.
[4] Falconer D, Einführung in die quantitative Genetik. Stuttgart: UTB; 1989
[5] Amos CI, Elston RC. Robust methods fort he detection of genetic linkage for quantitative data from pedigrees. Genet Epidemiol 1989; 6:349-360