

The COGA Study

COGA was a Collaborative study of the Genetics of Alcoholism, carried out at the Department of Psychiatry at Washington University, St. Louis, in the 1990s under the leadership of Professor Ted Reich. In the data set made available for study by other groups, including Genetic Analysis Workshop 11, there were 1214 individuals in 105 pedigrees (mostly 3 to 4 generations). The largest pedigrees were size 20 to 24, and only one had a loop (due to double first cousins), although there were many half-sibs.

There were genetic marker data on 992 of the 1214 individuals. There were 285 microsatellite markers on the 22 autosomes (average spacing 13.6 cM; this is quite wide). Accurate estimation of allele frequencies is important, because poor estimates can affect the analysis of the genetics of this quite complex trait.

The attached graphs by Tim Thornton show the impact of relatedness of individuals on the variance of the estimators of allele frequency, as a function of the true allele frequency q . The vertical axis says "MSE" or mean square error, but since we have unbiased estimators here this is the same as the variance.

Each curve is very close to $q(1-q)/n$ for some n – so we can think of n as the "effective" number of alleles in the sample. In fact n is about 375, 500, 515, 680 for the four curves.

The top two curves are for the naïve estimator – just count all the alleles, without taking relationships into account. These curves have large variance. The X-chromosome markers have larger variance in ratio 4:3 compared to markers on the autosomes. Why?

The lower two curves are for BLUE estimators – "best linear unbiased estimators". These are not MLEs --- MLEs could be quite hard to find and are likely not unbiased. However, the BLUEs are probably quite close to MLEs in these large samples. The important point is they take the relatedness among individuals into account. They have smaller variance. Again the ratio for the X to the autosomes is about 4:3.

The ratio of the naïve to the BLUE happens to be also about 4:3, for these particular pedigrees.