Genotyping error detection in samples of unrelated individuals without replicate genotyping.

OBJECTIVE: Identifying genotyping errors is an important issue in genetic research, yet it has been relatively less studied in samples consisting of unrelated individuals. In this article, we consider several models of genotyping errors, which were originally proposed for pedigree data, for unrelated population samples with single nucleotide polymorphism (SNP) genotype data. The mathematical constraints are investigated for detecting genotyping errors without resampling replicates or genotyping relatives.

METHODS: For the various proposed genotyping error models, we unveil the conditions under which the parameters are identifiable. These results are verified through applications to simulated and real SNP data.

RESULTS: We show that, with constraints, two particular models provide both identifiable error rate and allele frequencies of an SNP for unrelated population data. The simulation study shows that these two models present unbiased estimates for the allele frequencies. One of the models also gives an unbiased estimate for the genotyping error rate.

CONCLUSION: While the Hardy-Weinberg equilibrium test can be used to detect genotyping errors, a key advantage of these models is the explicit estimates of genotyping error rates and allele frequencies. This work may help researchers to estimate error rates and to use the estimates in their analysis to increase power and decrease bias, without the extra work of genotyping family members or replicates.