

“Missing Data in Case-Parent Triad Studies”

Tracy L. Bergemann,

Division of Biostatistics, School of Public Health,
University of Minnesota

A commonly used design in genetic association studies is the case-parent triad design. Generally, samples are drawn from an affected offspring, manifesting a disease or phenotype of interest, as well as from the parents. The trio genotypes may be analyzed using a variety of available methods, but we focus on log-linear models because they test for genetic association and additionally estimate the relative risks of transmission. The models need to be modified to impute missing genotypes. Furthermore, instability in the parameter estimates can arise when certain kinds of genotype combinations do not appear in the dataset.

In this research, we kill two birds with one stone. We propose a new method to simultaneously impute missing genotype data and account for genotype combinations with zero counts. This approach solves a zero-inflated Poisson (ZIP) regression likelihood. The maximum likelihood estimates yield relative risks and a likelihood ratio test determines the significance of genetic association.

We compared the ZIP regression approach to previously proposed methods in both simulation studies and a dataset investigating the risk of orofacial clefts. The ZIP likelihood estimates relative risks with less bias than other methods. Further, the new method preserves the appropriate type I error rate more carefully.