

Correcting for measurement errors in admixture estimates

Jasmin Divers¹, David T. Redden^{1,3}, José R. Fernandez^{1,2,3}, David A. Allison^{1,2,3}

¹Department of Biostatistics, Section on Statistical Genetics,

²Department of Nutrition Sciences

³Clinical Nutrition Research Center

University of Alabama at Birmingham

Birmingham, Alabama

An admixed population is created when two or more populations that have certain genes with disparate allele frequencies inter-mate. Because individuals from recently admixed populations exhibit large regions of genomic linkage disequilibrium, several researchers have suggested that linkage and association studies of complex phenotypes can be performed more efficiently in admixed populations. For such an approach to be successful, the proportion of an individual's genome that comes from a specific parental population must be estimated and controlled for statistically. Due to the interest in admixed populations, software has been developed that can provide these estimates. However, a critical and under researched issue for genetic studies seeking to control for admixture is the most statistically appropriate way to utilize an individual's admixture as a covariate. It is well documented that the introduction of a covariate measured with error leads to biased estimation of the parameters. Admixture estimates represents such a covariate. This problem, often referred to as a measurement error model, has been studied quite extensively in other research fields and the moment reconstruction method represents one of the newest proposed approaches to address it. Using simulations, we will i) compare the true admixture values estimated within simulated data to the admixture estimates provided by the software STRUCTURE, ii) apply three different measurement error correction methods: moment reconstruction, regression calibration and SIMEX to account for the measurement error in the admixture estimates, iii) compare these methods to see which one performs best, and iv) demonstrate that the parameters estimated using the reconstructed variables are less biased than those obtained ignoring the measurement errors.