Analysis of Population Structure and Stratification in NHANES III Self-Reported Race/Ethnicities

Christopher L. Sanders[1]
Ajay J. Yesupriya[2]
Lester R. Curtain[1]

[1] **National Center for Health Statistics, Centers for Disease Control and Prevention**
[2] **National Office of Public Health Genomics, Centers for Disease Control and Prevention**

Abstract

We assessed the population structure of 6,597 participants in a nationally representative population based on phase 2 of NHANES III using multilocus genetic data. Structure, a program that utilizes a Bayesian clustering approach, was used to infer the number of populations in our sample as well as estimate the probabilities that an individual's genetic background was from one or more of these populations. Analyses suggested that there were three distinct ancestral populations. Based on the self-reported race/ethnicity of our sample, we assumed these to be European, African, and Amerindian populations. Results indicated that individuals that self-reported non-Hispanic White and non-Hispanic Black were predominately of European and African ancestries, respectively. However, we found evidence of two distinct populations in self-reported Mexican-Americans that were primarily of European or Amerindian ancestry. Significant demographic differences between these two subpopulations suggest population stratification that could confound statistical analyses involving genetic markers in self-reported Mexican-Americans.

Keywords: Population, Structure, Stratification, NHANES, National Health and Nutrition Examination Survey, Self-reported, Race, Ethnicity