Mortality Selection in a Genetic Sample and Implications for Association Studies

Benjamin W. Domingue[1]
Daniel W. Belsky[2]
Amal Harrati[1]
Dalton Conley[3]
David Weir[4]
Jason Boardman[5]

Abstract

Mortality selection is a general concern in the social and health sciences. Recently, existing health and social science cohorts have begun to collect genomic data. Causes of selection into a genomic dataset can influence results from genomic analyses. Selective non-participation, which is specific to a particular study and its participants, has received attention in the literature. But mortality selection—the very general phenomenon that genomic data collected at a particular age represents selective participation by only the subset of birth cohort members who have survived to the time of data collection—has been largely ignored. Here we test the hypothesis that such mortality selection may significantly alter estimates in genetic association studies of both health and non-health traits. We demonstrate mortality selection into genome-wide SNP data collection at older ages using the US Health and Retirement Study (HRS). We then model the selection process. Finally, we test whether mortality selection alters estimates from genetic association studies. We find evidence for mortality selection. Healthier and more socioeconomically advantaged individuals are more likely to survive to be eligible to participate in the genetic sample of the HRS. Mortality selection leads to modest drift in estimating time-varying genetic effects, a drift that is enhanced when estimates are produced from data that has additional mortality selection. There is no general solution for correcting for mortality selection in a birth cohort prior to entry into a longitudinal study. We illustrate how genetic association studies using HRS data can adjust for mortality selection from study entry to time of genetic data collection by including probability weights that account for mortality selection. Mortality selection should be investigated more broadly in genetically informed samples from other cohort studies.