

Understanding genetic bias

Benjamin W. Domingue

Daniel Belsky

Jason D. Boardman

A classic problem in much of social science is to understand the causal relationship between an environment or behavior and some specific outcome. One example is the relationship between health and education (Adler & Rehkopf, 2008; Ross & Mirowsky, 1999). The Rubin Causal Model (e.g., Holland, 1986) offers a framework for understanding such relationships under certain conditions. Controlling for all relevant confounders is one relevant condition. In the example of health and education, neighborhood and school environments (Leventhal & BrooksGunn, 2003) or smoking (Farrell, Fuchs, and Fuchs, 1982) are confounders that could bias estimates of effects if left uncontrolled. Another confounder that could lead to bias in a variety of areas is genotype. If two traits, such as health and education, are at least partially a function of a common set of genes, then failing to account for this could lead to “genetic bias” in causal estimates. Our paper focuses on benchmarking the quantity of genetic bias that might exist under certain conditions and evaluation of various methods for removing genetic bias.