

Selected for Attainment: Estimating Early Child Health Effects on Adult Outcomes

Robert G. White¹

May 5, 2009

Abstract

Sample selection bias is a chronic problem in longitudinal studies that is particularly problematic for studies concerning the relationship between health and socio-economic status. This paper adopts two alternate methods for handling sample selection bias attributable to survey attrition and item non-response. Both methods are applied to examine the magnitude of bias in the effects of childhood cognition and behavior on the adult socio-economic gradient in health. A method for sample selection correction with multiple imputation for item non-response is implemented to account for different sources of sample selection bias over time. Estimates of a life course model of health and socioeconomic attainment demonstrate that sample selection bias inflates estimates of socioeconomic gradients. The proposed correction for sample selection bias also suggests that the effects of early child non-cognitive skills rather than cognitive skills may play an important role in the early life origins of adult socioeconomic gradients.

¹Center for Demography and Ecology, Department of Sociology, University of Wisconsin, Madison. rwhite@ssc.wisc.edu. ²Institute for Policy Research, Department of Sociology, Northwestern University.

I. Introduction

Socio-economic gradients in adult health outcomes have been among the most regularly observed empirical phenomena in the social sciences. From the 19th century (Chadwick 1842) through early studies of mortality and education (Kitagawa and Hauser 1973) to the growth in social epidemiology following the Whitehall II study (Marmot et al. 1991), alternate measures of good health over different periods of the life course consistently correspond with higher socio-economic status. Many explanations of these relationships update Kitagawa and Hauser's (1973) initial explanation of health gradients as reflecting differences in resources which accrue with educational attainment and that influence contemporaneous health. Such resources may include individual behaviors, such as smoking, drinking and body mass index (Kaplan et al. 1987; Preston and Taubman, 1994; Ecob and Davey Smith, 1999; Currie and Moretti, 2003) in addition to the effects of lower socioeconomic status on both health services utilization (Link and Phelan, 1995; Ross and Wu, 1995, Gornick et al., 1996; Kirby and Kaneda, 2006) and the duration of exposure to stress (Link and Phelan 1995; Stafford and Marmot, 2003; Dowd and Goldman 2006). An alternate set of explanations reverse the direction of causality and attribute adverse health with changes in wealth and labor market outcomes which lower trajectories of socioeconomic attainment (Shorrocks, 1975). Given the importance of the mid- and late-stages of lifetime labor market participation for career advancement and asset accumulation, health shocks in middle adulthood may pose particularly large risks for future socioeconomic status (Adams et al., 2002; Smith, 1999; Smith, 2003; Case and Deaton 2003).

While there is wide agreement over the importance of both sets of causal pathways, it is also increasingly evident that the origins of socioeconomic gradients in health precede many of these contemporaneous and near-term relationships. Growing attention across the medical and

social sciences to earlier periods in the life course have extended the time period for considering cumulative effects of low socioeconomic status and risky health behaviors to childhood.

Hypotheses tracing adult chronic conditions to prenatal and infant health (Barker et al. 1993; Barker 1995) have given rise to life course approaches in epidemiology (Kuh et al. 2003) that increasingly link birthweight and infant growth to adult risks for diabetes and cardiovascular disease. Evidence from twins studies (Conley et al. 2003; Behrman and Rosenzweig 2004; Black et al. 2007; Oreopoulos et al. 2008) and quasi-experimental methods (Almond 2006; Bleakley 2007; Shillingford 2008; Clark et al. 2008) also point to early health for similarly long lasting effects in schooling outcomes and socioeconomic attainment. The accumulating evidence from these approaches suggest dynamic ties between health and socioeconomic attainment over a wide span of the life course.

An important challenge for any study that addresses such a large number of hypothesized pathways over different periods of the life course is minimizing the loss of sample representativeness due to survey attrition and item nonresponse. These concerns are especially acute in longitudinal studies. Both item nonresponse and attrition have been associated with socioeconomic background, education, occupation, income (Groves and Couper 1988; Bound and Krueger 1991) and other characteristics that are among the main pathways among the hypothesized links between socio-economic attainment and health. Refreshment sampling in later survey rounds that over samples individuals along the predictors of subsequent nonresponse may only complicate the nature of this bias. In particular, without knowledge of the duration in low socio-economic status, changes in the effects of socio-economic status over time will result in further bias from over sampling low socio-economic status individuals. The directions of these biases and their changes over time are not easily predicted. Moreover, correcting such bias

is further complicated by changing probabilities of survey nonresponse over time that may also depend on socio-economic status.

This paper implements a procedure for simultaneously addressing item missingness and attrition. A conventional markov chain monte carlo procedure for multiple imputation is implemented in combination with a weighting method for sample selection correction. This procedure aims to relax the demands for multiple imputation by conditioning imputation on a weighting scheme that corrects selection bias related to attrition. The procedure is implemented in a life course model of early health and adult economic attainment using the National Child Development Study (NCDS) 1958 birth cohort from the United Kingdom. This dataset has a rich collection of measures for studying development, health and socioeconomic attainment over the life course and spans the life course from birth to age 46. A simulation exercise is then undertaken to assess the sensitivity of the results to the underlying assumptions of both the attrition correction method and multiple imputation.

The paper proceeds in section two by developing a method for sample selection correction that combines a weighting scheme and a MCMC method for multiple imputation. Section three proposes a model for linking early health to adult socioeconomic gradients in health that permits examining the consequences of the proposed method for life course studies of health and socioeconomic attainment. Section four describes the NCDS and illustrates the select process of survey participation that threatens unbiased estimates of life course models. Section five implements the proposed procedure and presents results. Section six presents the simulation results.