

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

Robert G. White

University of Wisconsin – Madison

September 20, 2008

Abstract

Sample selection bias is a chronic problem in longitudinal studies but one which is only rarely addressed in studies concerning ties between health and socio-economic status. This paper adopts two alternate methods for handling sample selection bias. Both methods are applied to examine the magnitude of bias in the effects of early health on educational attainment and adult socio-economic status. A sample selection model and a multiple imputation method for handling missing data are each adapted to account for different types of nonresponse and the possibility for changing severity in sample selection bias over time. Estimates from both methods are significantly larger than effects in the absence of any treatment for sample selection. The results demonstrate the importance of explicitly taking account of sample selection bias and suggest the possible magnitude of downward bias in standard estimates of socio-economic gradients in health.

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

Robert G. White
University of Wisconsin – Madison

September 20, 2008

Extended Abstract

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), numerous measures of good health over different periods of the life course consistently correspond with higher socio-economic status. While many of the early interpretations of socio-economic gradients in health outcomes emphasized the contemporaneous correlation between health and social status, increasing attention across the social sciences and epidemiology literatures to earlier periods in the life course raise the possible importance of early childhood events. Many of these analyses update Kitigawa and Hauser's (1973) initial explanation of health gradients as reflecting differences in behaviors which accrue with advances in educational attainment. While more recent explanations account for multiple possible pathways tying different attributes of early socio-economic status to later health outcomes, the most common emphasis in the sociology and social epidemiology literatures remains a causal relationship in which socio-economic status affects later health. In contrast, emerging research in economics identifies the lasting consequences of adverse early health conditions for both educational attainment and adult labor market outcomes. This work alternately examines health in infancy and middle adulthood and outcomes spanning cognition, academic achievement, career advancement and income. The range of these hypothesized pathways underscores the many nuanced dimensions in the ties

between health and socio-economic attainment over the life course. Given the large number and possibly conflicting nature of the policy recommendations which might follow from these pathways, it also raises the demands for stronger evidence of causality.

The widening use of 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) has made important contributions to reducing the uncertainty in estimates from models of socio-economic status and health estimated with survey research. Managing the effects of sample selection in survey research has not met comparable success. While advances in sampling methodology, the growing availability of survey frames and increasing innovation in recruitment techniques offer increasing alternatives for selecting appropriate samples for testing specific hypotheses and addressing particular public policies, sample attrition, survey nonresponse and missing data remain serious concerns across survey research. Sample selection arising from these three problems is especially acute in longitudinal studies and remains a serious concern in studies of socio-economic status and health.

Yet, the inadequate treatment of sample selection that is common in studies of socio-economic status and health raises questions about many of the central findings in this area of research. Ignoring sample selection has well known consequences for both the significance and magnitude of the central effects underlying health gradients. For instance, low socio-economic status is often a predictor of subsequent survey nonresponse, resulting in potentially large bias in estimates of socio-economic effects on later outcomes. 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.

Many methods exist for managing the effects of missing data when missingness occurs at random. However, individuals with selective item nonresponse have often been shown to share socio-demographic characteristics such as education, location and income (Groves and Couper, 1988; Bound and Krueger, 1991) that are common in studies of socio-economic attainment and health. Predicting the aggregate effects of these different sources of bias is difficult given the possible differences in the directions of the biases across different subpopulations as well as different reasons for survey nonresponse (e.g., death, refusals to participate, failure to locate respondents). Given the importance of hypotheses concerning attributes of socio-economic status, their relationship with health and the changes in any such relationships over time, the combination of these biases raises the possibility of severe distortion in model estimates.

The magnitude of this bias may also be especially pernicious in distinguishing the pathways between childhood events and later life health. The prevailing interpretations of adult health gradients emphasize the casual effects of either contemporaneous or prior socio-economic status on health. A large number of studies alternately attribute health services use (Ross and Wu, 1995; Gornick et al., 1996; Kirby and Kaneda, 2006), risky health behaviors (Kaplan et al., 1987; Ecob and Davey Smith, 1999; Currie and Moretti, 2003), social support (Berkman and Syme, 1979; Kennedy et al., 1998; Lochner et al., 2001) or the stress induced by relative social status (Link and Phelan, 1995; Marmot, 2001; Stafford and Marmot, 2003; Dowd and Goldman

2006; Link et al., 2008) as determinants of present or later life health. Yet, there is also evidence of the reverse relationship. Poor health in midlife has been linked to lost days of work, reduced savings and downward shifts in career trajectories (Shorrocks, 1975; Adams et al., 2002; Smith, 1999; Smith, 2005; Banks et al., 2006). Among the most common findings across these studies is the adverse consequences of poor early health for educational attainment (Wadsworth 1986; Currie and Hyson, 1999; Matte et al., 2001; Conley and Bennett 2000; Case et al. 2005). Moreover, the growth in epidemiologic evidence of ties between adverse childhood health and adult chronic conditions (Barker and Osmond, 1986; Barker et al., 1993; Barker, 1995; Leon et al. 1998; Blackwell et al., 2001) raises the possibility that any causal effects between health and socio-economic status may be limited to early childhood. Given the well known effects of parents' socio-economic status on lifetime socio-economic trajectories, it is then only necessary for family socio-economic status to affect health early in life to establish parallel yet unrelated life course trajectories toward low socio-economic status and poor health in adulthood.

This paper adopts two alternate methods for handling sample selection bias to examine the magnitude of bias in effects of early health on educational attainment and adult socio-economic status. The first method is a sample selection model that specifies the process of survey nonresponse (Heckman 1976, 1979). This approach includes estimating bounds for the main parameters of interest to provide additional estimates of the uncertainty due to sample selection (Manski 1989, 1990, 1995; Horowitz and Manski, 1998). The second approach adopts a bayesian method for multiple imputation that specifically addresses the effects of item nonresponse (Rubin 1976, 1987). Both approaches account for different types of nonresponse and the possibility for changes in their effects over the life course.

The two methods are applied using the National Child Development Study (NCDS). This longitudinal survey follows a sample of individuals born during a single week in 1958 in England, collecting extensive information about child health, development, educational attainment and ultimately, adult labor market activities. The NCDS remains one of the most complete longitudinal studies containing measures of child development and adult outcomes and consequently, has been the subject of many studies estimating the effects of early child health on adult outcomes. Combined with the long period of this study (seven waves) over many periods of the life course (birth through age 46), this cohort provides an ideal sample for assessing the scope of sample selection bias in models of socio-economic status and health over the life course. It has also been the subject of past efforts to identify the magnitude of sample selection bias. Plewis et al. (2004) describe the considerable efforts NCDS managers undertook to minimize survey nonresponse at all waves of the study and report analyses of the predictors of nonresponse that were used to inform refreshment sampling. Additional efforts to identify the severity of survey nonresponse also explicitly evaluated the magnitude of sample selection bias (Hawkes and Plewis, 2006). Both studies concluded that the evidence of nonrandom nonresponse could only minimally affect most studies of interest using the NCDS. These studies attribute the refreshment samples with mitigating any concerns about the effects of sample selection. These findings should provide reassurance to many users of the NCDS. However, they provide little insight into the possible severity of sample selection bias for life course studies that cover longer periods of time and which may be directly concerned with studying the measures which proved correlated with nonresponse. Moreover, the absence of any analysis of item nonresponse further leaves unanswered the possible effects both types of bias may have in model estimates of socio-economic status and health over the life course.

I find that both the severity of nonresponse and its predictors are highly variable over the length of the survey. Wave non-response is predicted for each wave controlling for observed respondent characteristics lagged over the period of the survey. These covariates are selected from a large set of covariates according to model fit statistics. Estimates for wave non-response are used to assess the extent to which non-response in the NCDS biases estimates of early child health effects on educational attainment and adult labor market status. It is shown that early child health, non-cognitive measures and socio-economic background predict survey non-response throughout the survey. However, the magnitudes of these effects change over the period of the survey and change in different directions amounting to potentially ambiguous aggregate effects depending on the ultimate model of interest.

While these effects may have uncertain aggregate effects for general models of socio-economic status and health, both methods for handling attrition result in larger estimates of the effects of early health effects on educational attainment and adult socio-economic status. The selection model specifies the probability of participating in the survey as a function of significant covariates identified in the attrition analysis. Estimates from the selection model result in larger magnitudes in the estimates of low birth weight effects on measures of cognitive and non-cognitive skills in adolescence and educational attainment. Estimates of the effects of cognitive and non-cognitive skills on both adult socio-economic status and adult self-reported health are also considerably larger than in the absence of any treatment of sample selection. Multiply imputing missing values for cases of item nonresponse results in only minor changes in the precision of the above model estimates.

While the results confirm the negative effects of low birth weight on cognitive skills and education which have been widely reported, the larger magnitudes of these effects when even

modest measures are adopted to handle sample selection underscore the possible underestimation of low birth weight effects on lifetime socio-economic attainment. These effects also demonstrate the important role that non-cognitive skills play in mediating the effects of birth weight on education. Moreover, they show how measures of both cognitive and non-cognitive skills during adolescence exert sustained effects on socio-economic status and health during adulthood. Combined, these findings outline a clear set of pathways linking early life health to later life socio-economic and health status which may play an important role in the persistence of adult socio-economic gradients in health.

References

- Adams, Peter, Michael D. Hurd, Daniel McFadden, Angela Merrill, and Tiago Ribeiro. 2003. "Healthy, Wealthy and Wise? Tests for Direct Causal Paths Between Health and Socioeconomic Status," *Journal of Econometrics* 112(1): 3-56.
- Almond, Douglas. 2006. "Is the 1918 Influenza Pandemic Over? Long-term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population" *Journal of Political Economy*, 114(4): 672-712.
- Banks, M., M. Marmot, Z. Oldfield, and J. P. Smith. 2006. "Disease and Disadvantage in the United States and in England." *Journal of the American Medical Association* 295:2037-45.
- Barker D.J.P. 1995. "Fetal origins of coronary heart disease." *British Medical Journal*, 311, 171-174.
- Barker, D.J., Osmond C. 1986. "Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales." *Lancet* 1:1077-81.
- Barker, D.J., Gluckman PD, Godfrey KM, Harding JE, Owens JA, Robinson JS. 1993. "Fetal nutrition and cardiovascular disease in adult life." *Lancet* 341: 938-41.
- Behrman Jere R., and Mark R. Rosenzweig. 2004. "Returns to Birth weight", *Review of Economics and Statistics*, May, 86(2), 586-601.
- Berkman, L.F. and S.L. Syme. 1979. "Social Networks, Host Resistance and Mortality: A 9-Year Follow-Up of Alameda County Residents." *American Journal of Epidemiology*, 109: 186-204.
- Black, Sandra E., Paul J. Devereux, and Kjells G. Salvanes. 2007. "From the cradle to the labor market? The effect of birth weight on adult outcomes," *Quarterly Journal of Economics*.
- Blackwell D.L., M.D. Hayward and E.M. Crimmins. (2001). "Does childhood health affect chronic morbidity in later life?" *Social Science and Medicine*, 52, 1269-1284.
- Bleakley, H. "Disease and Development: Evidence from Hookworm Eradication in the American South," *Quarterly Journal of Economics*, Feb. 2007.
- Bound, J. and A. Krueger, 1991. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics*, 9:1-24.
- Case, A., D. Lubotsky and C. Paxson. 2002. "Economic status and health in childhood: the origins of the gradient," *American Economic Review* 5:1308-34.
- Chadwick, E. 1842. *Report of an Enquiry into the Sanitary Conditions of the Laboring Population of Great Britain*. London: Poor Law Commission.
- Clarke, S. E., M. C Jukes, J. Njagi, L. Khasakhala, B. Cundill, J. Otido, C. Crudder, B. Estambale and S. Brooker. 2008. "Effect of intermittent preventive treatment of malaria on health and education in schoolchildren: a cluster-randomised, double-blind, placebo-controlled trial," *Lancet* 372: 127-138.
- Conley, D. and N. Bennett. 2000. "Is Biology destiny? Birth weight and life chances," *American Sociological Review* 65: 458-67.
- Conley, D., K. Strully, and N. Bennett. 2003. *The starting gate: birth weight and life chances*. Berkeley: University of California Press.
- Currie, J. and R. Hyson. 1999. "Is the impact of health shocks cushioned by socioeconomic status? The case of low birth weight," *American Economic Review* 2: 19-22.
- Currie, J. and E. Moretti. 2003. "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings," *Quarterly Journal of Economics*, 118(4): 1495-1532.
- Dowd, Jennifer Beam, and Noreen Goldman. 2006. "Do Biomarkers of Stress Mediate the Relation Between Socioeconomic Status and Health?" *Journal of Epidemiology and Community Health* 60: 633-639.
- Ecob, R., Davey Smith, G., 1999. Income and health: what is the nature of the relationship. *Social Science and Medicine* 48, 693-705.
- Gornick, M.E., P.W. Eggers, T.W. Reilly, et al. 1996. "Effects of race and income on mortality and use of services among Medicare beneficiaries." *New England Journal of Medicine* 335:791-799.
- Groves, Robert M. and Mick P. Couper. 1998. *Nonresponse in Household Interview Surveys*, New York: John Wiley and Sons.
- Hawkes, D. and I Plewis. 2006. "Modelling non-response in the National Child Development Study," *Journal of the Royal Statistical Society A* 169(3): 479-491
- Heckman, J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement* 5(4):475-492.
- Heckman, J. 1979. "Sample Selection Bias as a Specification Error," *Econometrica* 47(1):153-162.

- Horowitz, Joel L. and Charles F. Manski. 1998. "Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations," *Journal of Econometrics* 84(1):37–58.
- Kaplan, G. A., T. E. Seeman, R. D. Cohen, L. P. Knudsen, and J. Guralnik. 1987. "Mortality among the Elderly in the Alameda County Study: Behavioral and Demographic Risk Factors." *American Journal of Public Health* 77:307–12.
- Kennedy, B.P., I. Kawachi, R. Glass and D. Prothrow-Stith. 1998. "Income distribution, socioeconomic status, and self rated health in the United States: multilevel analysis" *British Medical Journal* 317(7163):917-21.
- Kitigawa, E., and P. Hauser. 1973. *Differential Mortality in the United States*. Cambridge: Harvard University Press.
- Kirby, J. B. and T. Kaneda, 2006. "Access to Health Care - Does Neighborhood Residential Instability Matter." *Journal of Health and Social Behavior*, 47: 142–155.
- Leon, D. A., H.O.Lithell, D. Vågero, I. Koupliova, R. Mohsen, L. Berglund, U. Lithell, and P. M. McKeigue. 1998. "Reduced Fetal Growth Rate and Increased Risk of Death from Ischaemic Heart Disease: Cohort Study of 1,500 Swedish Men and Women Born 1919–29." *British Medical Journal* 317(7153): 241–45.
- Link, B.G. and J. Phelan. 1995. "Social Conditions as Fundamental Causes of Disease," *Journal of Health and Social Behavior* 35: 80-94.
- Link, Bruce G., Jo C. Phelan, Richard Miech, and Emily Leckman. 2008. "The Resources that Matter: Fundamental Social Causes of Health Disparities and the Challenge of Intelligence." *Journal of Health and Social Behavior* 49(1):72-91.
- Lochner, K., E. Pamuk, D. Makuc, B.P. Kennedy, I. Kawachi. 2001. "State-Level Income Inequality and Individual Mortality Risk: A Prospective, Multilevel Study." *American Journal of Public Health* 91(3):385–91.
- Manski, Charles F. 1989. "Anatomy of the Selection Problem," *Journal of Human Resources* 24: 343–360.
- Manski, Charles F. 1990. "Nonparametric Bounds on Treatment Effects," *American Economic Review Papers and Proceedings* 80:319–323.
- Manski, Charles F. 1995. *Identification Problems in the Social Sciences* Cambridge, MA: Harvard University Press.
- Marmot, M. G. 2001. "Inequalities in Health." *New England Journal of Medicine* 345:134–36.
- Marmot, M. G., G.D. Smith, S. Stansfield, C. Patel, F. North, J. Head, I. White, E. Brunner, and A. Feeny. 1991. "Health Inequalities among British Civil Servants: The Whitehall II Study." *Lancet*, 337(8754): 1387–93.
- Matte, Thomas D., Michaeline Bresnahan, Mellisa D. Begg, and Ezra Susser. 2001. "Influence in Variation in Birth Weight within Normal Range and within Sibships on IQ at Age 7 Years: Cohort Study." *British Medical Journal* 323:310–14.
- Oreopoulos, Philip, Mark Stabile, Randy Walld, and Leslie L. Roos. 2008. "Short-, Medium-, and Long-Term Consequences of Poor Infant Health: An Analysis Using Siblings and Twins." *Journal of Human Resources* 43(1): 88–138.
- Plewis, I., Calderwood, L., Hawkes, D. and Nathan, G. 2004. "National Child Development Study and 1970 British Cohort Study Technical Report: Changes in the NCDS and BCS70 Populations and Samples over Time," London: Institute of Education.
- Ross, Catherine E. and Chia-Ling Wu. 1995. "The Links between Education and Health." *American Sociological Review* 60:719–45.
- Rubin, D. B. 1976. "Inference and missing data" *Biometrika* 63:581-592.
- Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys* New York: John Wiley and Sons.
- Shillingford, A.J., M. Glanzman, R.F. Ittenbach, R.R. Clancy, J.W. Gaynor and G. Wernovsky. 2008. "Inattention, Hyperactivity, and School Performance in a Population of School-Age Children With Complex Congenital Heart Disease," *Pediatrics* 121:e759-e767.
- Shorrocks, A. F. 1975. "The age–wealth relationship: a cross-section and cohort analysis," *Review of Economics and Statistics* 57:155–63.
- Smith, J.P. 1999. "Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Economic Status." *Journal of Economic Perspectives* 13(2):145-66.
- Smith, J. P. 2005. "The impact of SES on health over the life-course." Working Paper 318, RAND Labor and Population.
- Stafford, M., and M. Marmot. 2003. "Neighborhood Deprivation and Health: Does it Affect Us All Equally?" *International Journal of Epidemiology*, 32: 357–366.
- Wadsworth, M.E. 1986. "Serious Illness in Childhood and Its Association with Later-Life Achievement," in R. Wilkinson (ed.) *Class and Health: Research and Longitudinal Data*, London: Tavistock. pp 50-74.