Socioeconomic Status and Health across the Life Course: A Test of the Social Causation and Health Selection Hypotheses

John Robert Warren, *University of Minnesota*

This research investigates the merits of the "social causation" and "health selection" explanations for associations between socioeconomic status and self-reported overall health, musculoskeletal health and depression. Using data that include information about individuals' SES and health from childhood through late adulthood, I employ structural equation models that account for errors in measured variables and that allow for explicit tests of various hypotheses about how SES and health are related. For each outcome and for both women and men the results provide no support for the health selection hypothesis. SES affects each health outcome at multiple points in the life course, but the reverse is not true.

Inverse relationships between socioeconomic status and health have been documented for as long as public health researchers and others have been making observations (e.g., Chaplin 1924; Coombs 1941; Villerme 1840; Virchow 1848). In recent decades, researchers have pursued two related lines of inquiry in order to better understand relationships between SES and health. First, as reviewed below, they have investigated competing hypotheses about the nature of the causal relationship between SES and health. Does SES primarily affect health (the "social causation" hypothesis) or does health primarily affect SES (the "health selection" hypothesis)? Or, do they affect one another reciprocally or dynamically across the life course? And to what extent are observed associations between SES and health spurious owing to factors that subsequently influence both SES and health (the "indirect selection" hypothesis)? Second, researchers have investigated the mechanisms that intervene in SES-health relationships. That is, they have sought out the mediating variables through which SES operates to affect health and/or vise versa.

Versions of this article were presented at the 2006 annual meetings of the American Sociological Association and the Gerontological Society of America. This research uses data from the Wisconsin Longitudinal Study. Since 1991, the WLS has been supported principally by the National Institute on Aging (R01-AG-9775 and P01-AG-21079), with additional support from the Vilas Estate Trust, the National Science Foundation, the Spencer Foundation, and the Graduate School of the University of Wisconsin-Madison. *Direct correspondence to John Robert Warren, Department of Sociology, University of Minnesota, 909 Social Sciences, 267 19th Avenue South, Minneapolis, MN 55455. E-mail: warre046@umn.edu.*

This research provides information about the merits of the social causation and health selection hypotheses. Because the merits of these alternative hypotheses may vary across measures of SES (Alder and Ostrove 1999; Marmot et al. 1997) and across health outcomes (Dohrenwend et al. 1992; Huurre et al. 2005), I analyze the relationships between SES (as measured in a variety of ways) and three distinct health outcomes: self-assessed overall health, musculoskeletal health and depression. Furthermore, because relationships between SES and health unfold dynamically over the life course (Graham 2002), the analyses are based on data that include observations of individuals' SES and health from childhood through late adulthood, and utilize structural equation models that allow for explicit tests of the various hypotheses about how SES and health are related. Finally, because relationships between SES and health may differ for women and men (Dohrenwend et al. 1992; Elliott 2001; Muntaner et al. 2004), separate models are estimated for women and men; the analyses include tests of hypotheses about sex differences in the nature of SES-health relationships.

Social Causation and Health Selection

The social causation hypothesis – which stipulates that SES affects health through mechanisms that are not entirely understood – has been well-developed by social scientists and public health researchers (Nielsen, Juon and Ensminger 2004; Wheaton 1978) and has received empirical support in a variety of contexts (e.g., Grundy and Sloggett 2003; Power et al. 2002). Grundy and Sloggett (2003:935), for example, found that "socio-economic indicators, particularly receipt of income support (a marker of poverty) were… consistently associated with raised odds of poor health outcomes."

On the other hand, the health selection (or drift) hypothesis stipulates that individuals' health influences their ability to attain or maintain desirable socioeconomic positions and resources (Blane, Davey Smith and Bartley 1993; Eaton 1980; Elstad 2001; Perrot and Collins 1935; Power, Matthews and Manor 1996; West 1991). For example, using longitudinal data collected in Hagerstown, Maryland, Laurence (1958) concluded that most of the observed associations between SES and chronic illness could be attributed to the tendency for chronically ill individuals to drift downward in their social and economic circumstances. More recently, Chandola et al. (2002) found significant effects of mental health on changes in SES among men in the Whitehall II study.

While the social causation and health selection hypotheses differ with respect to the primary *direction* of the causal relationship between SES and health, the "indirect selection" hypothesis stipulates that the relationship between SES and health is – at least to some extent –

spurious owing to factors that influence both SES and health. For example, there is considerable evidence that fetal environment, birth weight and childhood health influence both adult health and educational attainment – an important indicator of SES itself as well as a key determinant of other aspects of SES such as income and occupational attainment (Case, Fertig and Paxson 2005; Case, Lubotsky and Paxson 2002; Currie and Stabile 2004; Kuh and Wadsworth 1993). Consequently, relationships between SES and health as observed in adulthood may reflect – at least in part – the impact of childhood health (or SES for that matter) on adult SES and health. In order to assess the relative merits of the social causation and health selection hypotheses, it is important to account statistically for indirect selection.

Debating the relative merits of the social causation and health selection hypotheses has occupied researchers for decades (Bartley and Plewis 1997; Blane, Davey Smith and Bartley 1993; Dohrenwend et al. 1992; Elstad and Krokstad 2003; Fox 1990; Illsley 1986; Miech et al. 1999; West 1991; Wheaton 1978). Although analysts have most carefully focused on the direction of causality in relationships between SES and *mental* health outcomes (Dohrenwend et al. 1992; Miech et al. 1999; Muntaner et al. 2004; Power et al. 2002; Wheaton 1978), there is also a growing literature focusing on a variety of indicators of *physical* health and of *self-assessed overall health* (Cardanoa, Costab and Demaria 2004; Chandola et al. 2003; Elstad and Krokstad 2003; Manor, Matthews and Power 2003; Mulatu and Schooler 2002). Overall the empirical evidence regarding the relative value of the social causation and health selection hypotheses is mixed, varying by the way in which SES and health are measured and by the methodology utilized. As noted recently by Huurre et al. (2005:581), "[t]he present state of understanding regarding the origins of health disparities suggests that social class differences in health cannot be explained by a single model. Several processes of both causation and selection may be at work, the relative importance varying from one condition to another."

Some recent evidence, however, has cast doubt on the utility of the health selection hypothesis in a number of contexts (Dohrenwend et al. 1992; Power et al. 2002). Using longitudinal data from the Whitehall II study, Chandola and colleagues (2003:2059) concluded that "the development of social gradients in [mental and physical] health in the Whitehall II study may not be primarily explained in terms of a health selection effect." Likewise Manor, Matthews and Power (2003:2226) used data from the 1958 British cohort and found that "health selection has a variable effect on the social gradient in health, which is modest in size and cannot be regarded as a major explanation for social inequalities in health." Although some research is oriented toward assessing which one of the two hypotheses is "correct," more thoughtful analyses recognize that both may have merit and that the

goal is to understand whether social causation or health selection play greater or lesser relative roles in particular circumstances. For example Mulatu and Schooler (2002:22) found that "both social causation and health selection contribute to social inequalities in health" in their sample of American men and women.

Methodological Considerations

Differences in methodology – as well as differences in how SES and health are conceptualized and measured – likely account for many of the discrepant findings about the relative contributions of social causation and health selection in generating inequalities in health outcomes. Three issues in particular – accounting for indirect selection, estimating models of reciprocal or lagged causation, and accounting for errors in measured variables – are particularly relevant in the present context.

First, without sufficient controls for factors that influence both SES and health it is impossible to determine for sure whether and to what extent SES and health influence one another. Analysts sometimes include measures of neither childhood SES nor childhood health (Chandola et al. 2003; Mulatu and Schooler 2002), although they more frequently include measures of childhood SES but not of childhood health (Elstad and Krokstad 2003; Wheaton 1978). The challenge, of course, is to obtain long-term longitudinal data for a representative sample of respondents who are followed for a long period of time beginning in adolescence. Even analysts with access to data on both childhood SES and childhood health have typically been limited to observing relatively short periods of respondents' lives (e.g., Manor, Matthews and Power 2003; Miech et al. 1999; Power et al. 2002). The present analyses utilize data on SES and health for a representative sample of respondents who have been followed from adolescence through retirement age.

Second, while researchers routinely recognize the complex interplay between SES and health across the life course, they rarely utilize analytic techniques that allow them to model that complexity. Surprisingly few researchers have made use of structural equation modeling techniques to estimate cross-lagged models of the effects of SES and health at one point in time on their counterparts at a later point in time (Chandola et al. 2003; Huurre et al. 2005). Even fewer have estimated models of reciprocal causation in which SES and health are hypothesized to affect one another simultaneously (see Mulatu and Schooler 2002 for an exception). The present analyses utilize structural equation modeling techniques that allow SES and health to affect one another dynamically across the life course and that provide for explicit tests of the social causation and health selection hypotheses.

Third, although the deleterious consequences of measurement error for regression results has long been acknowledged in both the social science and public health literatures (Armstrong 1990; Bielby, Hauser and Featheman 1977; Hauser, Tsai and Sewell 1983; Thomas, Stram and Dwyer 1993), only Wheaton (1978), Mulatu and Schooler (2002), and a few others attempt to account for errors in measured variables in their analyses of the relative merits of the social causation and health selection hypotheses. In a standard OLS regression model, measurement error in the endogenous variable does not bias the estimated regression slopes (although it does lead to less efficient estimates). On the other hand, measurement error in exogenous variables does bias estimated regression parameters. Because the present analyses must recognize the dynamic interplay of SES and health across the life course, and because measurement error may thus have serious consequences for key regression estimates, the present analyses utilize structural equation modeling techniques that account for errors in measured SES and health variables.

Many previous researchers have tested the relative merits of the social causation and health selection hypotheses. The current investigation makes a number of contributions to the literature. First, the data at my disposal are unique in a number of respects: Respondents are followed prospectively for some 50 years; SES is measured in childhood and at two points in adulthood; health is measured at two points in adulthood and retrospective information about childhood health is obtained in adulthood; and the presence of multiple indicators of all measures allows for methodological advances over most prior research. Second, the structural equation models allow for a more nuanced assessment of the complex interplay between SES and health across the life course. Surprisingly few researchers have estimated empirical models that correspond to their conceptual notions of how SES and health affect one another across the life course.

Data and Measures

The Wisconsin Longitudinal Study is a long-term study of a random sample of 10,317 men and women who graduated from Wisconsin high schools in 1957.[1] Most respondents were born in 1939 and thus slightly predate the baby boom generation. The WLS provides an opportunity to study the life course, intergenerational transfers and relationships, family functioning, physical and mental health and well-being, and morbidity and mortality from late adolescence through middle age. WLS data also cover social background, youthful aspirations, schooling, military service, labor market experiences, family characteristics and events, social participation, psychological characteristics and retirement. Survey data were collected from the original respondents or their parents by mail and/or telephone

in 1957, 1964, 1975, 1993 and 2004; from a selected sibling in 1977, 1994 and 2005; from the spouse of the original respondent in 2004; from the spouse of the selected sibling in 2005; and from widow(er)s of the graduates and siblings in 2005. A new round of in-person and mail-back surveys of graduates, siblings and spouses will commence in 2009-2010.

The WLS has unique strengths, including the combination of an unusually wide array of high-quality measures of health and socioeconomic status as collected across the life course for a large sample that is broadly representative of non-Hispanic white high school graduates and that has been followed (with high rates of sample retention, as described below) for a half a century. Despite these unique strengths, the WLS data also have obvious limitations. Some strata of American society are not represented. Everyone in the primary sample graduated from high school; there are few black, Hispanic or Asian persons in the WLS; about 19 percent grew up on a farm; and in each post-1957 survey wave about 70 percent of participants has lived in Wisconsin. Although this means that the WLS cohort is not representative of the U.S. population in general, data from the 1960 U.S. Census indicate that at that time about 60 percent of Americans in the WLS birth cohort (1937 to 1940) were white high school graduates; about 75 percent of Wisconsin residents in this birth cohort were white high school graduates in 1960 (Sewell and Hauser 1975). The fact that the WLS cohort is not representative of the entire U.S. population is clearly a limitation, but the cohort is representative of a large segment of the U.S. population.

My analyses primarily utilize data from five surveys of the main sample of WLS graduate respondents: the 1975 telephone survey, the 1993 telephone and mail surveys, and the 2004 telephone and mail surveys. I begin by restricting my analytic sample to the 5,300 individuals who responded to each of the telephone and mail surveys in 1975, 1993 and 2004. I also exclude from the sample 10 respondents who were not born in the years 1937-1940, yielding a final sample size of 5,290: 2,394 men, 2,896 women. This is 51 percent of the original 10,317 respondents, but 59 percent of those who are known to have survived to 2004. Appendix A provides a more complete description of the pattern and nature of non-response across surveys. Among other things, it shows that the final sample of 5,290 – which includes only 51 percent of the original 10,317 panelists – is disproportionately female due to differential response and mortality rates, and modestly over-represents individuals from more advantaged socioeconomic origins. As shown in the appendix, a fair number of those excluded from the sample are deceased; the remainder did not respond to one or more of the several surveys. All of the measures used in my analyses – along with the number of observations for which data are available – are described in Table 1. As shown in the table, there is a small amount of missing data on individual items. Because there is

relatively little missing data, all of the structural equation models described below are estimated using pairwise-present covariance matrices.

Measures of childhood socioeconomic circumstances were obtained in 1957 and/or 1964 and in some cases were later updated or supplemented in 1975 and/or 1993. Father's and mother's education is expressed in years of schooling completed. Family income in 1957 is expressed in logged 1957 dollars and was derived from Wisconsin state tax records. Father's occupation in 1957 is coded to the standards of Duncan's (1961) Socioeconomic Index. I also include indicators of whether respondents lived with both parents while growing up, how many siblings respondents had, and whether respondents were from farm backgrounds (as determined by their fathers' occupation).

In 2004 respondents were asked a series of retrospective questions about their health when they were growing up (through age 16). First, they were asked, "Was your health as a child excellent, very good, good, fair or poor?" Next, respondents were asked to indicate whether they had any of the following illnesses or treatments as a child or young adult: Asthma, frequent ear infections, removal of tonsils and/or adenoids, chronic bronchitis, whooping cough, polio, diphtheria, hepatitis, pneumonia, meningitis or infectious mononucleosis. The present analyses utilize a scale that indicates how many of these illnesses/treatments respondents had; as shown in Table 1, the typical respondent had about one of these illnesses or treatments. Finally, respondents were asked questions about whether health conditions up through age 16 caused them to miss school for a month or more, confined them to home or bed for a month or more, or limited their sports or physical activities for three months or more. Table 1 shows that about 13 percent of respondents were limited in at least one of these ways up through age 16.

Respondents' educational attainment is expressed in terms of years of schooling completed, and was obtained in both the 1964 and 1975 surveys. Socioeconomic status in 1975, 1993 and 2004 is expressed in terms of log earnings, occupational socioeconomic standing, and family net worth (except that net worth was not available in 1975). The 1975 measure of log earnings pertains to total family earnings, whereas the 1993 and 2004 measures pertain to respondents' own hourly wage rates. In each year occupational socioeconomic standing is expressed as occupational education – the proportion of occupational incumbents who have completed at least some college (Hauser and Warren 1997). The 1993 and 2004 net worth measures – which are expressed in thousands of dollars – are based on a thorough accounting of respondents' family assets and debts.

The present analyses utilize measures of three different health outcomes: Self-reported overall health, musculoskeletal health, and depression. In both 1993 and 2004 respondents' self-reported overall

health is measured in three ways. First, in both years respondents were asked, "How would you rate your health at the present time?" Second, in both years respondents were asked, "How would you rate your health compared with other people your age and sex?" Response options for both questions included "very poor," "poor," "fair," "good" and "excellent." Entirely subjective measures of self-assessed overall health are closely linked to more objective measures of morbidity and mortality, and have

Table 1: Descriptive Statistics

	Men (n = 2,394)			Women (n = 2,896)		
	Avg or %	(SD)	n	Avg or %	(SD)	n
SES of Childhood Family						
Father's Education (in Years)	9.8	(3.5)	2,394	9.8	(3.4)	2,896
Mother's Education (in Years)	10.6	(2.8)	2,394	10.4	(2.9)	2,896
Family Income in 1957 (logged)	8.5	(.7)	2,092	8.5	(.7)	2,569
Father's Occupation (SEI)	33.1	(21.7)	2,373	33.4	(21.5)	2,877
Two Parent Family? (Yes = 1)	90.4%		2,392	90.7%		2,894
Number of Siblings	3.2	(2.5)	2,392	3.3	(2.5)	2,894
Farm Origin? (Yes=1)	18.2%		2,394	17.9%		2,896
Health in Childhood						
Overall Health (1 = Poor, 5 = Excellent)	4.3	(.8)	2,383	4.3	(.9)	2,890
Number of Serious Illnesses	1.0	(1.0)	2,394	1.1	(1.0)	2,896
Health Limitations? (Yes = 1)	12.8%		2,394	11.8%		2,896
Educational Attainment						
Years of Schooling Completed (1964)	13.8	(2.1)	2,172	13.3	(1.7)	2,596
Years of Schooling Completed (1975)	14.0	(2.5)	2,394	13.1	(1.9)	2,896
SES in 1975						
Total Family Earnings in 1974 (logged)	9.3	(2.0)	2,386	8.8	(2.7)	2,894
Occupation in 1975 (Occupational Education)	41.1	(30.8)	2,377	35.8	(27.8)	2,105
SES in 1993						
Hourly Wage Rate in 1993 (logged)	2.8	(.9)	2,305	2.0	(1.3)	2,609
Occupation in 1993 (Occupational Education)	62.4	(25.5)	2,385	62.0	(22.8)	2,703
Family Net Worth in 1993 ($1000s)	254.6	(239.4)	2,283	206.7	(213.5)	2,725
SES in 2005						
Hourly Wage Rate in 2005 (logged)	3.1	(1.0)	2,104	2.4	(1.2)	2,170
Occupation in 2005 (Occupational Education)	58.1	(25.8)	1,408	57.8	(22.7)	1,473
Family Net Worth in 2005 ($1000s)	516.0	(348.1)	2,394	380.0	(326.5)	2,895

been widely validated (Idler and Benyamini 1997; Wilson 2001). Third, in 2004 respondents were asked, "Including what you have already told me, would you say that you have EVER had any long-term physical or mental conditions, illnesses or disabilities that limited what you were able to do, either on or off the job?" Respondents who answered affirmatively were then asked about the nature of the two most serious conditions – which were then coded to the standards of the International Classification of

Health in 1993							
Overall Health in 1993 (1 = Poor, 5 = Excellent)	4.17	(.64)	2,389	4.19	(.65)	2,894	
Overall Health Rel. to Peers in 1993	4.21	(.69)	2,358	4.18	(.72)	2,865	
Has Any Serious Condition (Yes = 1)	5.8%		2,394	5.7%		2,896	
Back Pain or Strain in 1993	.08	(.22)	2,338	.10	(.26)	2,814	
Serious Back Trouble in 1993	.05	(.17)	2,394	.05	(.18)	2,894	
Has Any Musculoskeletal Condition in 1993	3.1%		2,394	4.0%		2,896	
CES-D Depression Score in 1993	14.8	(13.6)	2,370	16.9	(15.9)	2,873	
Health in 2005							
Overall Health in 2005 (1 = Poor, 5 = Excellent)	4.01	(.69)	2,386	4.02	(.70)	2,877	
Overall Health Rel. to Peers in 2005	4.13	(.75)	2,347	4.12	(.74)	2,836	
Has Any Serious Condition (Yes = 1)	13.3%		2,394	15.5%		2,896	
Back Pain or Strain in 2005	.12	(.26)	2,296	.16	(.31)	2,787	
Serious Back Trouble in 2005	.07	(.21)	2,330	.08	(.23)	2,803	
Has Any Musculoskeletal Condition in 2005	8.4%		2,394	10.6%		2,896	
CES-D Depression Score in 2005	12.2	(12.6)	2,342	14.5	(14.3)	2,846	

Note: See text for description of sample selection and variable construction.

Diseases (or ICD-9; Centers for Disease Control 2001) – and about the years in which those conditions began and ended limiting what they were able to do. Based on these dates, respondents were classified according to whether or not they were limited by any such condition in 1993 and in 2004.

Musculoskeletal health in 1993 and 2004 were each measured in three ways. First, in the 1993 and 2004 mail surveys respondents were asked, "In the past six months have you had back pain or strain?" Respondents who answered affirmatively were then asked, "How often have you experienced back pain or strain?" and "How much discomfort has back pain or strain caused you in the past six months?" Responses to these three items were combined into a single scale such that zero represents a respondent who experienced no back pain or strain and 1 represents a respondent who experienced "a lot" of discomfort from back pain or strain at least daily. Second, in the 1993 and 2004 mail surveys respondents were asked to respond to the statement, "A medical professional says you have serious back trouble." This time, respondents who responded affirmatively were

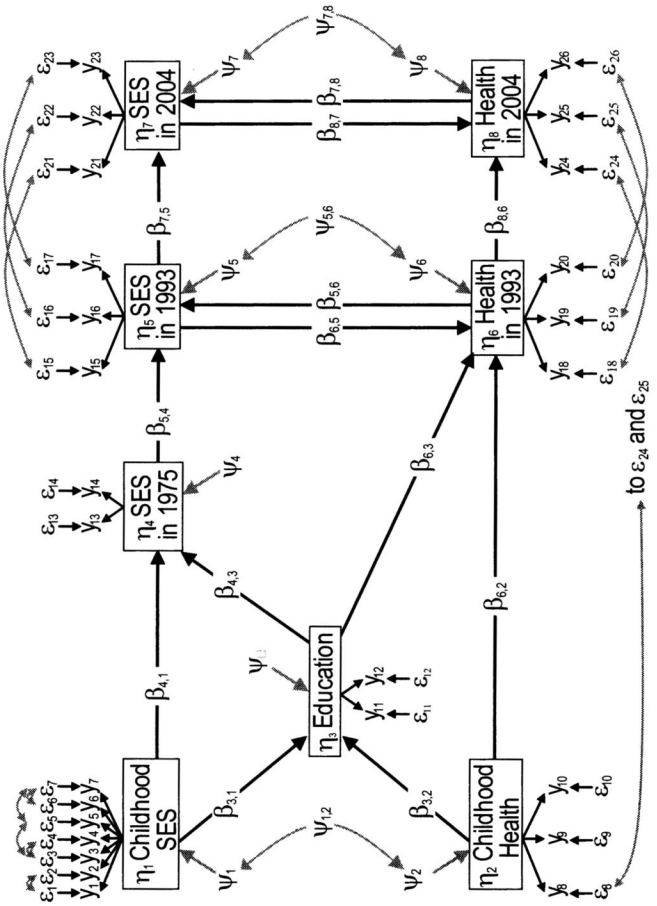


Figure 1: Baseline Model

asked, "How much does your serious back trouble currently interfere with what you like to do?" Responses to these two items were combined into a single scale such that zero represents a respondent who has not been diagnosed with serious back trouble and 1 represents a respondent who has been diagnosed with serious back trouble which interferes "a great deal" with what they like to do. Third, using the ICD-9 codes for the most serious conditions that limit what respondents are able to do, I classify respondents in 1993 and 2004 as having a musculoskeletal health condition if the ICD-9 code for their condition falls into the category, "Diseases of the Musculoskeletal System and Connective Tissue."

Finally, in the 1993 and 2004 WLS mail surveys, respondents were presented with 20 items that make up the CES-D depression scale. These include such items as, "On how many days during the past week did you feel you could not shake off the blues even with help from your family and friends?" As depicted in Table 1, the mean and variability of these scale scores declined modestly for both men and women from 1993 through 2004.

Research Design

Figure 1 depicts the baseline structural equation model that was estimated separately for each of the three health outcomes. The model specifies that each of the observed variables, y, is a function of both a latent variable, η, and a random disturbance, ϵ; Appendix Table B1 reports the loadings of each observed variable (y) on its corresponding latent variable (η) as well as estimated error variances (ϵ). As described above and in Table 1, the model includes seven indicators of childhood SES (η_1), three indicators of childhood health (η_2), two indicators of education (η_3), two indicators of SES in 1975 (η_4), three indicators each of SES in 1993 and 2004 (η_5 and η_7, respectively), and – in the case of overall health and musculoskeletal health – three indicators of health in 1993 and 2004 (η_6 and η_8, respectively); for depression, there is only one observed indicator in both 1993 and 2004.

For each latent variable, η, the loading of η on one of the several y variables is fixed to equal 1.0; consequently, that latent variable is in the metric of that y variable. Table 1 indicates – through the use of italics – the y variable whose factor loading was so constrained. Because the variables in Table 1 represent a mix of continuous and ordinal variables, all structural equation models are based on analyses of polychoric correlation matrixes. Consequently, all of the structural parameters in the model are standardized. Finally, all models were estimated separately for women and men; whether particular parameters are invariant across genders is of substantive interest, and setting up the structural equation models in a two-group format allows formal tests of the equality of parameters for women and men.

The baseline model specifies that childhood SES and childhood health are associated with one another and affect education. Childhood SES and education both directly affect SES in 1975, while childhood health and education both directly affect health in 1993. SES in 1975 affects SES in 1993, which in turn affects SES in 2004. Health in 1993 affects health in 2004. SES and health in 1993 affect one another reciprocally; likewise, SES and health in 2004 affect one another reciprocally. Substantively, two aspects of the model are most interesting. First, the model allows an assessment of the effects of SES at one point in time on health at that same point in time, and vise versa. As described above, researchers have only rarely tested the social causation and health selection hypotheses using models of reciprocal causation.[2] Second, the coefficient representing the effects of childhood health on education is substantively interesting. Even if health at one point in time does not directly affect SES at that point in time, if childhood health affects education – which in turn likely affects subsequent SES – then this provides support for the health selection hypothesis.

Results

Table 2 presents fit statistics for the baseline model and for alternately specified models, separately by health outcome. Decisions about improvements in model fit are based on changes in BIC (Raftery 1995); a reduction in BIC of 6 or more is viewed as reason to prefer one model over an alternative (nested or non-nested) model to which it is being compared. Model 2 imposes the constraint that the disturbances in the structural model – the Ψ parameters – are equivalent for men and women (whereas all other parameter matrixes are free to differ for men and women). For each outcome, the preferred model includes these constraints. Model 3 imposes the constraint that the structural coefficients in the model – the β parameters – are equivalent for men and women. For overall health and depression the preferred model includes this constraint. That is, for overall health and depression these results indicate no differences between men and women in the structural portion of the model; for musculoskeletal health men and women differ with respect to the β parameters. Model 4 begins with the preferred model for each health outcome – Model 3 for overall health and depression, Model 2 for musculoskeletal health – and imposes the constraints that (1. the effect of SES on health in 1993 is the same as the effect of SES on health in 2004 and that (2. the effect of health on SES in 1993 is the same as the effect of health on SES in 2004. For all three health outcomes the preferred model includes this set of constraints.

Models 5 and 6 begin with Model 4 and impose constraints that are designed to directly test the social causation and health selection hypotheses. Model 5 specifies that health in 1993 does *not* affect SES in

Table 2: Model Fit Statistics

	$\chi 2$	df	BIC
Overall Health			
Model 1: Baseline Model; No Equality Constraints by Sex	5855.9	588	1115
Model 2: Model 1 + Ψ Matrix Equivalent for Men and Women	5914.3	599	1085
Model 3: Model 2 + β Matrix Equivalent for Men and Women	5971.2	612	1037
Model 4: Model 3 + $\beta_{56} = \beta_{78}$ and $\beta_{65} = \beta_{87}$	5975.1	614	1025
Model 5: Model 4 but $\beta_{56} = 0$ and $\beta_{78} = 0$ (No Health Selection)	**5975.2**	**615**	**1017**
Model 6: Model 4 but $\beta_{65} = 0$ and $\beta_{87} = 0$ (No Social Causation)	6004.5	615	1046
Musculoskeletal Health			
Model 1: Baseline Model; No Equality Constraints by Sex	9135.9	588	4395
Model 2: Model 1 + ψ Matrix Equivalent for Men and Women	9207.6	599	4378
Model 3: Model 2 + β Matrix Equivalent for Men and Women	9374.5	612	4440
Model 4: Model 2 + $\beta_{56} = \beta_{78}$ and $\beta_{65} = \beta_{87}$	9234.9	603	4373
Model 5: Model 4 but $\beta_{56} = 0$ and $\beta_{78} = 0$ (No Health Selection)	**9252.5**	**605**	**4375**
Model 6: Model 4 but $\beta_{65} = 0$ and $\beta_{87} = 0$ (No Social Causation)	9299.7	605	4422
Depression			
Model 1: Baseline Model; No Equality Constraints by Sex	2847.4	418	-523
Model 2: Model 1 + Ψ Matrix Equivalent for Men and Women	2882.2	429	-577
Model 3: Model 2 + β Matrix Equivalent for Men and Women	2935.7	442	-628
Model 4: Model 3 + $\beta_{56} = \beta_{78}$ and $\beta_{65} = \beta_{87}$	2939.9	444	-640
Model 5: Model 4 but $\beta_{56} = 0$ and $\beta_{78} = 0$ (No Health Selection)	**2939.9**	**445**	**-648**
Model 6: Model 4 but $\beta_{65} = 0$ and $\beta_{87} = 0$ (No Social Causation)	2961.0	445	-627

Note: See text for description of sample selection, variable construction and model specification.

1993 and that health in 2004 does *not* affect SES in 2004; that is, Model 5 specifies that there is no health selection. Conversely, Model 6 specifies that there is no (direct) social causation by specifying that SES in 1993 does *not* affect health in 1993 and that SES in 2004 does *not* affect health in 2004. As shown in Table 2, for each health outcome Model 5 is preferred over Model 4; in no case is Model 6 preferred over Model 4. In other words, the preferred specification for each health outcome stipulates that health in 1993 has no effect on SES in 1993 and that health in 2004 has no effect on SES in 2004.

Tables 3 through 5 present the structural coefficients – the β parameters – for women and men for overall health, for musculoskeletal health, and for depression, respectively; the results of the measurement portion of the model are presented in Appendix Table B1. As specified above, these parameters are equivalent for men and women for overall health

and depression but not for musculoskeletal health. In all three sets of models higher childhood SES leads to greater educational attainment; better childhood health leads to fewer health problems in 1993; more educational attainment leads to higher SES in 1975; higher SES in 1975 leads to higher SES in 1993; higher SES in 1993 leads to fewer health problems in 1993 and to higher SES in 2004; fewer health problems in 1993 leads to fewer health problems in 2004; and higher SES in 2004 leads to fewer health problems in 2004. In no case does childhood health directly and significantly affect education, further weakening the case for the health selection hypothesis.

Discussion

A careful understanding of the nature of associations between SES and health – that is, an understanding of whether those associations are causal in nature and in which direction causality flows – is necessary before public health researchers and others can successfully elaborate the processes and mechanisms through which SES and health affect one another. The goal of this research was to assess the extent to which "social causation" (the hypothesis that SES causally affects health) and "health selection" (the hypothesis that health affects SES) account for observed associations between SES (measured in a variety of ways) and three very different health outcomes: self-assessed overall health, musculoskeletal health and depression.

Any analyses of the relative merits of the social causation and health selection hypotheses face a number of data and methodological issues. Like some previous work, the present analyses make use of indicators of SES and health as measured at multiple points in the life course, including in adolescence. They also employ structural equation modeling techniques to carefully model the dynamics of the relationships between SES and health across the life course. Also, the structural equation models account for errors in measured variables. The present analyses are unique in their combination of these several data and methodological strengths. What is more the present analyses are among the first that have considered such a long span of respondents' lives – from adolescence through retirement age – and are among the first to carefully test hypotheses about how social causation and health selection processes might operate differently for men and women.

Nonetheless, the present analyses are not without limitations. The analyses do not assess the impact of panel attrition or selective mortality on the empirical results. The fact that the childhood health measures are obtained retrospectively – at age 65 or 66 – is less than ideal, as is the fact that all of the health measures are self-reported. We have no measures of things such as fetal environment or birth weight, and these may be important omissions. There is currently no way to remedy the latter shortcoming

Table 3: Structural Coefficients in Preferred Model for Overall Health

	Exogenous Variable							
Endogenous Variable	Childhood SES	Childhood Health	Education	SES in 1975	SES in 1993	Health in 1993	SES in 2004	Health in 2004
Men and Women								
Childhood SES	—	—	—	—	—	—	—	—
Childhood Health	—	—	—	—	—	—	—	—
Education	**.29 (18.95)**	.03 (1.04)	—	—	—	—	—	—
SES in 1975	**-.05 (-5.57)**	—	**.80 (53.77)**	—	—	—	—	—
SES in 1993	—	—	—	**.88 (44.77)**	—	—	—	—
Health in 1993	—	**.38 (13.28)**	**.07 (2.78)**	—	**.12 (5.65)**	—	—	—
SES in 2005	—	—	—	—	**.80 (36.90)**	—	—	—
Health in 2005	—	—	—	—	—	**.81 (28.57)**	**.12 (5.65)**	—

Note: See text for description of sample selection, variable construction, and model specification. Numbers in parentheses are t-statistics; bolded coefficients are significant at the .05 level.

Table 4: Structural Coefficients in Preferred Model for Musculoskeletal Health

	Exogenous Variable							
Endogenous Variable	Childhood SES	Childhood Health	Education	SES in 1975	SES in 1993	Health in 1993	SES in 2004	Health in 2004
Men								
Childhood SES	—	—	—	—	—	—	—	—
Childhood Health	—	—	—	—	—	—	—	—
Education	**.31** **(14.38)**	.03 (.80)	—	—	—	—	—	—
SES in 1975	-.01 (-.89)	—	**.81** **(41.32)**	—	—	—	—	—
SES in 1993	—	—	—	**.91** **(37.43)**	—	—	—	—
Health in 1993	—	**-.17** **(-8.30)**	**.15** **(10.29)**	—	**-.16** **(-8.44)**	—	—	—
SES in 2005	—	—	—	—	**.77** **(28.95)**	—	—	—
Health in 2005	—	—	—	—	—	**.90** **(19.36)**	**-.16** **(-8.44)**	—
Women								
Childhood SES	—	—	—	—	—	—	—	—
Childhood Health	—	—	—	—	—	—	—	—
Education	**.26** **(13.50)**	-.03 (-.73)	—	—	—	—	—	—
SES in 1975	**-.08** **(-7.36)**	—	**.77** **(35.48)**	—	—	—	—	—

SES in 1993		—		—	**.85 (27.18)**		—		—		—	
Health in 1993		**-.28 (-9.56)**		**.08 (3.10)**	—		-.04 (-1.86)		—		—	
SES in 2005		—		—	—		**.85 (23.70)**		—		—	
Health in 2005		—		—	—		—		**1.52 (19.84)**		-.04 (-1.86)	

Note: See text for description of sample selection, variable construction, and model specification. Numbers in parentheses are t-statistics; bolded coefficients are significant at the .05 level.

using the WLS data. However, no other data include the wealth of high-quality information about SES and health across the life course for a representative sample of Americans, followed (with high rates of sample retention) for half a century.

The results are unambiguous. The hypothesis of social causation is strongly supported in the case of each health outcome and for both women and men. SES in 1993 has strong and significant effects on each health outcome in 1993; the same is true in 2004. The magnitude of the pertinent (standardized) coefficients indicates that a one standard deviation increase in SES leads to about a one-seventh standard deviation improvement in each health outcome in each year. In contrast, the empirical analyses provide no support for the health selection hypothesis for any of the three very different health outcomes. Models that stipulate no effect of health in 1993 on SES in 1993 and no effect of health in 2004 on SES in that year are preferred over models that allow for such effects. In models that do allow for SES and health to affect one another simultaneously, the coefficients (not shown) representing the effects of health in one year on SES in that year are rarely statistically significant. Although there is no evidence that health in adulthood affects SES in adulthood, the health selection hypothesis might still have been supported if childhood health had been shown to affect educational attainment. If this were the case, then I might have concluded that "social drift" begins early in the life course, such that childhood health affects adult SES indirectly, through its effect on education. However, in none of the preferred models does childhood health have statistically significant effects on educational attainment for either women or men.

Previous researchers have noted quite correctly that the relative merits of the

Table 5: Structural Coefficients in Preferred Model for Depression

	Exogenous Variable							
Endogenous Variable	Childhood SES	Childhood Health	Education	SES in 1975	SES in 1993	Health in 1993	SES in 2004	Health in 2004
Men and Women								
Childhood SES	—	—	—	—	—	—	—	—
Childhood Health	—	—	—	—	—	—	—	—
Education	**.71** **(25.39)**	.01 (.48)	—	—	—	—	—	—
SES in 1975	**-.08** **(-3.61)**	—	**.81** **(47.25)**	—	—	—	—	—
SES in 1993	—	—	—	**.88** **(44.94)**	—	—	—	—
Health in 1993	—	**-.24** **(-6.98)**	.03 (1.06)	—	**-.11** **(-4.65)**	—	—	—
SES in 2005	—	—	—	—	**.80** **(37.01)**	—	—	—
Health in 2005	—	—	—	—	—	**.53** **(35.61)**	**-.11** **(-4.65)**	—

Note: See text for description of sample selection, variable construction and model specification. Numbers in parentheses are t-statistics; bolded coefficients are significant at the .05 level.

social causation and health selection hypotheses may vary depending on the health outcome under consideration. While this is theoretically quite sensible, the present analyses cast doubt on the notion that health selection plays an important role for any particular health outcome. In models that address indirect selection by controlling for an array of measures of childhood SES and childhood health, that allow SES and health to affect one another dynamically across the life course, that correct for errors in measured variables, and that do not necessarily force results to be equivalent for women and men, there is no evidence that health selection accounts for observed associations between SES and any of these very different three health outcomes.

Notes

1. A public use file of data from the Wisconsin Longitudinal Study is available from the Data and Program Library Service, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, Wisconsin 53706 and at http://dpls.dacc.wisc.edu/WLS/wlsarch.htm

2. As shown in Figure 1, the model stipulates that SES and health affect one another simultaneously in both 1993 and 2004. Alternate models that instead allow for cross-lagged effects – for example, of SES in 1975 on health in 1993 and of SES in 1993 on health in 2004 – yield substantively similar results.

References

Alder, Nancy E., and Joan M. Ostrove. 1999. "Socioeconomic Status and Health: What We Know and What We Don't." Pp. 96-115. *Socioeconomic Status and Health in Industrialized Nations. Annals of the New York Academy of Sciences*, *Volume 896*. Nancy Adler, Michael Marmot, Bruce McEwen and Judith Stewart, editors. New York Academy of Sciences.

Armstrong, Ben G. 1990. "The Effects of Measurement Errors on Relative Risk Regressions." *American Journal of Epidemiology* 132(6):1176-84.

Bartley, Mel, and Ian Plewis. 1997. "Does Health-Selective Mobility Account for Socioeconomic Differences in Health? Evidence from England and Wales, 1971 to 1991." *Journal of Health and Social Behavior* 38(4):376-86.

Bielby, William T., Robert M. Hauser and David L. Featheman. 1977. "Response Errors of Black and Nonblack Males in Models of the Intergenerational Transmission of Socioeconomic Status." *American Journal of Sociology* 82(6):1242-88.

Blane, David, George Davey Smith and Mel Bartley. 1993. "Social Selection: What Does it Contribute to Social Class Differences in Health?" *Sociology of Health and Illness* 15(1):1-15.

Cardanoa, Mario, Giuseppe Costab and Moreno Demaria. 2004. "Social Mobility and Health in the Turin Longitudinal Study." *Social Science & Medicine* 58(8):1563-74.

Case, Anne, Darren Lubotsky and Christina Paxson. 2002. "Socioeconomic Status and Health in Childhood: The Origins of the Gradient." *American Economic Review* 92(5):1308-34.

Case, Anne, Angela Fertig and Christina Paxson. 2005. "The Lasting Impact of Childhood Health and Circumstance." *Journal of Health Economics* 24(2):365-89.

Centers for Disease Control. 2001. *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)*. Department of Health and Human Services Publication PHS 00-1260. Hyattsville, MD: National Center for Health Statistics.

Chandola, Tarani, Mel Bartley, Amanada Sacker, Crispin Jenkinson and Michael Marmot. 2003. "Health Selection in the Whitehall II Study, UK." *Social Science & Medicine* 56(10):2059-72.

Chaplin, Charles V. 1924. "Death Among Taxpayers and Non-Taxpayers, Providence, 1865." *American Journal of Public Health* 14(8):647-51.

Coombs, L.C. 1941. "Economic Differentials in Causes of Death." *Medical Care* 1:246-55.

Currie, Janet, and Mark Stabile. 2004. "Socioeconomic Status and Health: Why is the Relationship Stronger for Older Children?" *American Economic Review* 93(5):1813-23.

Dohrenwend, Bruce, Itzhak Levav, Patrick Shrout, Sharon Schwartz, Guedalia Naveh, Bruce Link, Andrew Skodol and Ann Stueve. 1992. "Socioeconomic Status and Psychiatric Disorders: The Causation-Selection Issue." *Science* 255(5047):946-52.

Duncan, Otis D. 1961. "A Socioeconomic Index for All Occupations." Pp. 109-38. *Occupations and Social Status*. Albert J. Reiss Jr., editor. Free Press.

Eaton, William W. 1980. "A Formal Theory of Selection for Schizophrenia." *American Journal of Sociology* 86(1):149-58.

Elliott, Marta. 2001. "Gender Differences in Causes of Depression." *Women Health* 33(3):163-77.

Elstad, Jon I., and Steinar Krokstad. 2003. "Social Causation, Health-Selective Mobility, and the Reproduction of Socioeconomic Health Inequalities Over Time: Panel Study of Adult Men." *Social Science & Medicine* 57(8):1475-89.

Elstad, Jon I. 2001. "Health-Related Mobility, Health Inequalities and Gradient Constraint." *European Journal of Public Health* 11(2):135-40.

Fox, John W. 1990. "Social Class, Mental Illness, and Social Mobility: The Social Selection-Drift Hypothesis for Serious Mental Illness." *Journal of Health and Social Behavior* 31(4):344-53.

Graham, Hilary. 2002. "Building an Inter-Disciplinary Science of Health Inequalities: The Example of Life Course Research." *Social Science & Medicine* 55(11):2005-16.

Grundy, Emily, and Andy Sloggett. 2003. "Health Inequalities in the Older Population: The Role of Personal Capital, Social Resources and Socioeconomic Circumstances." *Social Science & Medicine* 56(5):935-47.

Hauser, Robert M., Shu-Ling Tsai and William H. Sewell. 1983. "A Model of Stratification with Response Error in Social and Psychological Variables." *Sociology of Education* 56(1):20-46.

Hauser, Robert M., and John Robert Warren. 1997. "Socioeconomic Indexes of Occupational Status: A Review, Update, and Critique." Pp. 177-298. *Sociological Methodology 1997*. Adrian Raftery, editor. Blackwell Publishers.

Huurre, Taina, Ossi Rahkonen, Erkki Komulainen and Hillevi Aro. 2005. "Socioeconomic Status as a Cause and Consequence of Psychosomatic Symptoms from Adolescence to Adulthood." *Social Psychiatry and Psychiatric Epidemiology* 40(7):580-87.

Idler, Ellen, and Yael Benyamini. 1997. "Self-rated health and mortality: A review of twenty-seven community studies." *Journal of Health and Social Behavior* 38(1):21-37.

Illsley, R. 1986. "Occupational Class, Selection and the Production of Inequalities in Health." *Quarterly Journal of Social Affairs* 2(2):151-65.

Kuh, Diana, and Michael Wadsworth. 1993. "Physical Health Status at 36 Years in a British National Birth Cohort." *Social Science & Medicine* 37(7):905-16.

Laurence, P.S. 1958. "Chronic Illness and Socioeconomic Status." Pp. 39-49. *Patients, Physicians, and Illness*. E.Gartly Jaco, editor. Free Press.

Manor, Orly, Sharon Matthews and Chris Power. 2003. "Health Selection: the Role of Inter- and Intra-Generational Mobility on Social Inequalities in Health." *Social Science & Medicine* 57(11):2217-27.

Marmot, Michael, Carol D. Ryff, Larry L. Bumpass, Martin Shipley and Nadine F. Marks. 1997. "Social inequalities in health: Next questions and converging evidence." *Social Science & Medicine* 44(6):901-10.

Miech, Richard A., Avshalom Caspi, Terrie E. Moffitt, Bradley R.E. Wright and Phil A. Silva. 1999. "Low Socioeconomic Status and Mental Disorders: A Longitudinal Study of Selection and Causation During Young Adulthood." *American Journal of Sociology* 104(4):1096-1131.

Mulatu, Mesfin S., and Carmi Schooler. 2002. "Causal connections between socio-economic status and health: Reciprocal effects and mediating mechanisms." *Journal of Health and Social Behavior* 43(1):22-41.

Muntaner, Charles, William W. Eaton, Richard Miech and Patricia O'Campo. 2004. "Socioeconomic Position and Major Mental Disorders." *Epidemiologic Reviews* 26(1):53-62.

Nielsen, Marcia J., Hee-Soon Juon and Margaret Ensminger. 2004. "Preventing Long-Term Welfare Receipt: The Theoretical Relationship between Health and Poverty over the Early Life Course." *Social Science & Medicine* 59(11):2285-2301.

Perrot, George S., and Selwyn D. Collins. 1935. "Relation of Sickness to Income and Income Change in 10 Surveyed Communities. Health and Depression Studies No. 1. Method of Study and General Results for Each Locality." *Public Health Reports* 50:592-622.

Power, Chris, Sharon Matthews and Orly Manor. 1996. "Inequalities in Self Rated Health in the 1958 Birth Cohort: Lifetime Social Circumstances or Social Mobility?" *British Medical Journal* 313(7055):449-53.

Power, Chris, Stephan Stansfeld, Sharon Matthews, Orly Manor and S. Hope. 2002. "Childhood and adulthood risk factors for socio-economic differentials in psychological distress: evidence from the 1958 British birth cohort." *Social Science & Medicine* 55(11):1989-2004.

Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." Pp. 111-63. *Sociological Methodology*. P. Marsden, editor. Blackwell.

Sewell, William H., and Robert M. Hauser. 1975. *Education, Occupation, and Earnings: Achievement in the Early Career*. Academic Press.

Thomas, Duncan, Daniel Stram and James Dwyer. 1993. "Exposure Measurement Error: Influence on Exposure-Disease Relationships and Methods of Correction." *Annual Review of Public Health* 14:69-93.

Villerme, Louis 1840. *Tableau d'Etat Physique et Moral des Ouvriers, Volume 2*. Renouard.

Virchow, Rudolf. 1848. "The Public Health Service (in German)." *Medizinische Reform* 5:21-22.

West, Patrick 1991. "Rethinking the Health Selection Explanation for Health Inequalities." *Social Science & Medicine* 32(4):373-84.

Wheaton, Blair. 1978. "The Sociogenesis of Psychological Disorder: Reexamining the Causal Issues with Longitudinal Data." *American Sociological Review* 43(3):383-403.

Wilson, Sven E. 2001. "Socioeconomic Status and the Prevalence of Health Problems Among Married Couples in Late Midlife." *American Journal of Public Health* 91(1):131-35.

Appendix A. Description of WLS Sample

The WLS began as a survey of all Wisconsin high school seniors in 1957. A randomly selected 10,317 (about a third) of those seniors were selected for follow-up in 1964, and have constituted the "graduate" sample of the WLS ever since. Major follow-up surveys were conducted in 1975 (by telephone), in 1993 (by telephone and mail), and in 2004 (by telephone and mail). Details of the content of each survey and of fielding procedures can be found at http://www.ssc.wisc.edu/wlsresearch/.

An historic strength of the WLS has been its high rates of sample retention. In 1975, 90 percent of living respondents completed the telephone survey. In 1993, 87 percent of living respondents completed the telephone survey. Follow-up mail surveys in were sent to those who completed the 1993 telephone survey; 81 percent of those who were sent mail surveys completed them. In 2004, 80 percent of living respondents completed the telephone survey. In 2004, follow-up mail surveys were sent to all living sample members, regardless of whether they completed the 2004 telephone survey. At that time, 76 percent of living respondents completed the mail survey. In short, all of these major survey efforts have achieved at least 76 percent response rates, and most have achieved at least 80 percent response rates. Few major surveys can claim such high response rates, especially over such a long period of time.

However, the analyses presented in this article draw on data collected in each of five separate survey operations: the 1975 telephone survey, the 1993 telephone survey, the 1993 mail survey, the 2004 telephone survey, and the 2004 mail survey. While each of these five surveys enjoyed high response rates, my analyses can be conducted only on the subset of cases in which respondents participated in all five, for a total of 5,290 individuals.

In Figure A1 (for the full sample) and in Table A1 (for the full sample and separately by sex) I describe how the initial sample of 10,317 is reduced to a subset of 5,290 for my purposes. First, 9,139 individuals completed the 1975 telephone survey. Of these 9,139, only 6,548 completed *both* the telephone and mail surveys in 1993. Of these 6,548, only 5,300 completed *both* the telephone and mail surveys in 2004. As described in the text, 10 respondents were excluded from the final sample because they were not in the modal birth cohort. The final result is 5,290 cases: 2,394 men and 2,896 women. This is 5,290/10,317 = 51.2 percent of the original respondents. However, since 1,292 WLS sample members are known to have died by 2004, I might say that my final sample includes 5,290/(10,317-1,292) = 58.6 percent of all living cohort members. By sex, my final analysis sample includes 56.7 percent of all surviving men and 60.3 percent of all surviving women.

Table A1: Non-Response and Mortality, by Gender

	TOTAL		MEN		WOMEN	
		%		%		%
1957 In-School Survey						
Respondents	10,317	—	4,991	—	5,326	—
1975 Telephone Survey						
Respondents	9,139	88.6	4,330	86.8	4,809	90.3
Non-Respondents	1,004	9.7	543	10.9	461	8.7
Deceased	174	1.7	118	2.4	56	1.1
Total	10,317	100	4,991	100	5,326	100
1993 Telephone and Mail Surveys						
Respondents	6,548	71.6	3,031	70.0	3,517	73.1
Non-Respondents	2,250	24.6	1,110	25.6	1,140	23.7
Deceased	341	3.7	189	4.4	152	3.2
Total	9,139	100	4,330	100	4,809	100
2004 Telephone and Mail Surveys						
Respondents	5,300	80.9	2,399	79.1	2,901	82.5
Non-Respondents	861	13.1	412	13.6	449	12.8
Deceased	387	5.9	220	7.3	167	4.7
Total	6,548	100	3,031	100	3,517	100
Born 1937-1940	5,290		2,394		2,896	

Note: 1975 respondents include those who participated in the 1975 telephone survey; 1993 respondents include those who participated in the 1993 telephone and mail surveys; and 2004 respondents include those who participated in the 2004 telephone and mail surveys. At each wave, non-respondents include individuals who were not located and individuals who were located but who were unwilling to participate. Of the 5,300 respondents in 2004, 10 were dropped from the analyses because they were not in the modal birth cohort.

The 5,290 who are included in my analysis sample are obviously not a strictly random subset of the original 10,317. Table A1 makes clear, for example, that response rates are higher among women and mortality rates are higher among men. In Table A2 I compare respondents to non-respondents and to deceased cohort members with respect to sex, educational attainment, and socioeconomic origins – characteristics that can be measured using pre-1975 survey data. The comparisons for 1993 are made only among those who responded in 1975. The comparisons for 2004 are made only among those who responded in 1993. This setup facilitates analyses of differential non-response and differential mortality between each survey administration. Table A2 demonstrates that respondents are disproportionately female, both because of differential response rates and differential mortality. Table A2 also shows that respondents are better educated than either non-respondents or deceased panelists. Although the differences by education are statistically significant, they are small in

magnitude. Respondents are better educated than either non-respondents or deceased panel members, but only by a fraction of a year. Respondents also tend to be from slightly more advantaged socioeconomic origins – as measured by father's education and family income – but even when statistically significant these differences are small in magnitude.

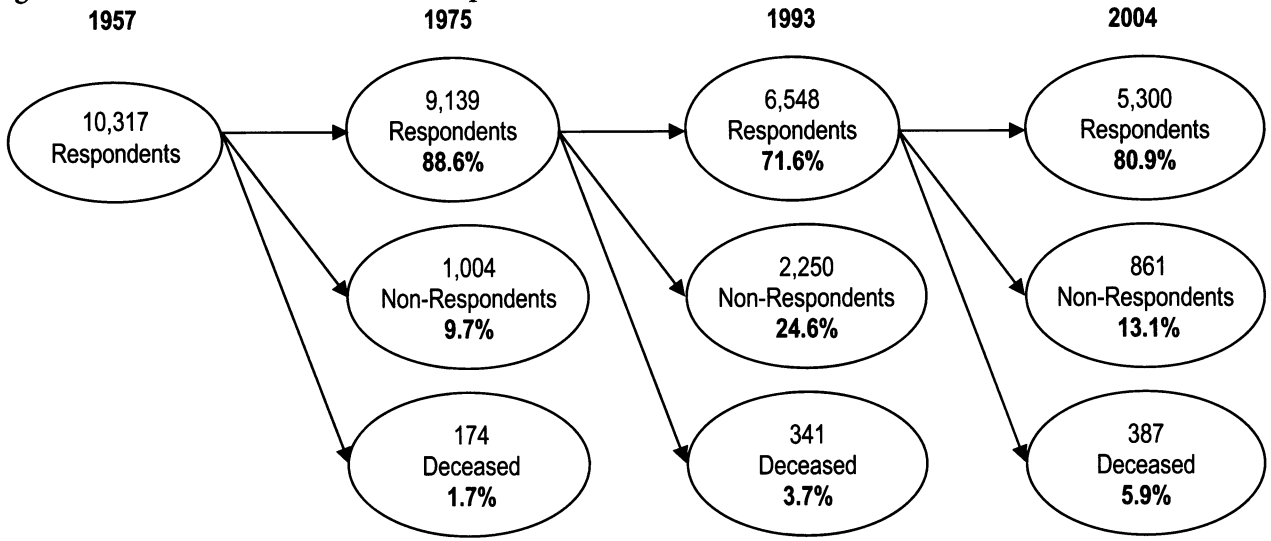
My analysis sample includes 59 percent of WLS graduates who survived to 2004. Although not ideal, this figure is fairly high given that graduates had to have completed each of five major surveys across a span of several decades in order to be included in the analyses. Individuals included in the analysis sample are disproportionately female, and – because of survivorship bias – are almost certainly disproportionately healthy. They also tend to be modestly better educated and to be from slightly more advantaged socioeconomic origins. As described in the text, the loss of 41 percent of surviving panelists and the non-random pattern of selection into the final sample are clearly limitations of the present analyses.

Table A2: Gender, Education, and Family SES Differences between Respondents, Non-Respondents and Deceased Sample Members

	Respondents		Non-Respondents			Deceased		
	Avg or %	(sd)	Avg or %	(sd)		Avg or %	(sd)	
1975 Telephone Survey	(n = 9,139)		(n = 1,004)			(n = 174)		
Male	47.4%		54.1%		**	67.8%		**
Years of Schooling Completed (in Years)	13.4	(1.8)	13.1	(1.7)	**	13.2	(1.6)	n.s.
Father's Education (in Years)	9.7	(3.4)	9.8	(3.0)	n.s.	9.8	(2.9)	n.s.
Family Income in 1957 (logged)	8.51	(.70)	8.44	(.75)	**	8.30	(.77)	**
1993 Telephone and Mail Surveys	(n = 6,548)		(n = 2,250)			(n = 341)		
Male	46.3%		49.3%		**	55.4%		**
Years of Schooling Completed (in Years)	13.5	(1.9)	13.1	(1.6)	**	13.3	(1.7)	*
Father's Education (in Years)	9.8	(3.4)	9.6	(3.3)	*	9.7	(3.5)	n.s.
Family Income in 1957 (logged)	8.53	(.69)	8.48	(.70)	**	8.45	(.78)	n.s.
2004 Telephone and Mail Surveys	(n = 5,300)		(n = 861)			(n = 387)		
Male	45.3%		47.9%		n.s.	56.8%		**
Years of Schooling Completed (in Years)	13.5	(1.9)	13.2	(1.7)	**	13.3	(1.8)	**
Father's Education (in Years)	9.8	(3.5)	9.6	(3.2)	n.s.	9.7	(3.4)	n.s.
Family Income in 1957 (logged)	8.54	(.68)	8.51	(.76)	n.s.	8.51	(.65)	n.s.

Notes: Hypothesis tests compare non-respondents to respondents and, separately, deceased sample members to respondents. n.s. = Not Significant **$p < .01$ *$p < .05$

Figure A1. Attrition from the WLS Sample across Waves



Note: 1975 respondents include those who participated in the 1975 telephone survey; 1993 respondents include those who participated in the 1993 telephone and mail surveys; and 2004 respondents include those who participated in the 2004 telephone and mail surveys. At each wave, non-respondents include individuals who were not located and individuals who were located but who were unwilling to participate. Of the 5,300 respondents in 2004, 10 were dropped from the analyses because they were not in the modal birth cohort.

Table B1: Factor Loadings and Error Variances for Measurement Components of Best-Fitting Models, by Health Outcome and Sex

	Men (n = 2,394)						Women (n = 2,896)					
	Overall Health		Musculoskeletal Health		Depression		Overall Health		Musculoskeletal Health		Depression	
	λ_y	ϵ	λ_y	ϵ	λ_y	ϵ	λ_y	ϵ	λ_y	ϵ	λ_y	ϵ
SES of Childhood Family												
Father's Education (in Years)	.50	.73	.51	.73	1.04	.44	.40	.80	.38	.81	1.02	.46
Mother's Education (in Years)	.33	.88	.33	.88	.79	.68	.36	.84	.34	.85	.93	.55
Family Income in 1957 (logged)	.66	.51	.66	.51	.69	.75	.76	.36	.77	.34	.71	.74
Father's Occupation (SEI)	1.00	-.07	1.00	-.07	1.00	.46	1.00	-.21	1.00	-.24	1.00	.51
Two Parent Family? (Yes = 1)	.08	.99	.08	.99	.07	.99	.07	.99	.07	.99	.03	.99
Number of Siblings	-.16	.97	-.16	.97	-.33	.94	-.14	.98	-.13	.98	-.38	.92
Farm Origin? (Ye s= 1)	-.64	.57	-.64	.57	-.61	.77	-.58	.59	-.56	.60	-.57	.83
Health in Childhood												
Overall Health (1 = Poor, 5 = Excellent)	1.00	.47	1.00	.52	1.00	.59	1.00	.45	1.00	.54	1.00	.56
Number of Serious Illnesses	-.66	.77	-.74	.75	-.79	.74	-.59	.81	-.65	.80	-.70	.79
Health Limitations? (Yes = 1)	-.78	.68	-.85	.66	-.99	.58	-.88	.58	-1.03	.50	-1.08	.50
Educational Attainment												
Years of Schooling Completed (1964)	1.00	.14	1.00	.12	1.00	.13	1.00	.12	1.00	.13	1.00	.12
Years of Schooling Completed (1975)	1.05	.04	1.03	.06	1.04	.04	1.05	.05	1.05	.04	1.04	.05
SES in 1975												
Total Family Earnings in 1974 (logged)	.07	.99	.07	.99	.07	.99	.10	.99	.11	.99	.10	.99
Occupation in 1975 (Occupation Education)	1.00	.25	1.00	.26	1.00	.24	1.00	.32	1.00	.33	1.00	.32
SES in 1993												
Hourly Wage Rate in 1993 (logged)	.48	.86	.48	.86	.46	.86	.38	.91	.40	.92	.37	.91
Occupation in 1993 (Occupation Education)	1.00	.32	1.00	.35	1.00	.30	1.00	.45	1.00	.52	1.00	.43
Family Net Worth in 1993 (1000s)	.39	.91	.39	.91	.38	.91	.33	.94	.31	.95	.31	.94
SES in 2005												
Hourly Wage Rate in 2005 (logged)	.45	.89	.49	.87	.44	.89	.44	.90	.44	.91	.43	.90

Occupation in 2005 (Occupation Education)	1.00	.46	1.00	.49	1.00	.44	1.00	.52	1.00	.54	1.00	.51
Family Net Worth in 2005 (1000s)	.61	.80	.67	.78	.59	.81	.50	.88	.45	.91	.49	.88
Health in 1993												
Overall Health in 1993 (1 = Poor, 5 = Excellent)	1.00	.34		—		—	1.00	.34		—		—
Overall Health Rel. to Peers in 1993	.91	.46		—		—	1.05	.27		—		—
Has Any Serious Condition (Yes = 1)	-.76	.63		—		—	-.68	.69		—		—
Back Pain or Strain in 1993		—	1.00	.70		—		—	1.00	.77		—
Serious Back Trouble in 1993		—	.80	.77		—		—	1.33	.57		—
Has Any Musculoskeletal Condition in 1993		—	1.86	.15		—		—	1.12	.65		—
CES-D Depression Score in 1993		—		—	1.00	.00		—		—	1.00	.00
Health in 2005												
Overall Health in 2005 (1 = Poor, 5 = Excellent)	1.00	.17		—		—	1.00	.24		—		—
Overall Health Rel. to Peers in 2005	.90	.33		—		—	1.02	.20		—		—
Has Any Serious Condition (Yes = 1)	-.61	.71		—		—	-.73	.60		—		—
Back Pain or Strain in 2005		—	1.00	.54		—		—	1.00	.22		—
Serious Back Trouble in 2005		—	1.07	.40		—		—	1.12	.01		—
Has Any Musculoskeletal Condition in 2005		—	1.11	.50		—		—	.58	.72		—
CES-D Depression Score in 2005		—		—	1.00	.00		—		—	1.00	.00

Note: See text for description of sample selection and variable construction.