

NOTES AND COMMENTS

[Slide 1] **Healthy, Wealthy, and Wise? The Causal Pathways between Health and Socioeconomic Status.** Daniel McFadden, University of California, Berkeley, and NBER, June 16, 2003.

[Slide 2] The links between health and socioeconomic status have been studied in a number of populations, with the general finding that higher SES is associated with better health and longer life. Surveying this literature, Nadine Goldman (2001) notes that this association has been found in different eras, places, genders, and ages, and occurs over the whole range of SES levels.

[Slide 3] The association holds for a variety of health variables (most illnesses, mortality, self-rated health status, psychological well-being, and biomarkers such as allostatic load) and alternative measures of SES (wealth, education, occupation, income, level of social integration).

[Slide 4] This table gives the relative risk of low versus high SES in a sample of elderly in the U.S. in 1993. I will say more later about the sample. Relative risks that are statistically significant at the five percent level are highlighted. People are low SES if they are in the bottom quartile for income and wealth, have less than a high school education, and live in a dwelling in poor condition in a distressed neighborhood. They are high SES if they are in the top quartile for income and wealth, have a college education, and live in a dwelling in good condition in a good neighborhood. The abbreviations are for high blood pressure (HBP), cognitive impairment (Cog. Imp.), and poor/fair self-reported health status (P/F SRHS). These relative risks are calculated holding constant factors such as concurrent health conditions and smoking history. Taken at face value, these results suggest that low SES is a major risk factor for many health conditions.

[Slide 5] The ubiquity of the association between health and wealth suggests that it cannot be explained solely by poverty, social discrimination, the absence of basic medical technology in LDCs, selective availability of advanced technology in developed nations, or different systems for delivery and financing of health care. Rather, the root causes must be the heterogeneities in nature and nurture that are present in all human populations. This does not mean that health and economic policies have no causal role, or are harmless or ineffective in influencing the association and redressing inequities. It does mean that careful dissection is needed to separate out the policy-relevant direct causal effects from the ecological correlations induced by population heterogeneities. The econometrician's mantra is "Correlation and coincidence do not imply causation." A story makes the point. A father was telling his son about great baseball players. He said Babe Ruth was probably the greatest ballplayer and next was probably Lou Gehrig, who may even have been the greatest but his career was cut short by a terrible disease. The son asked "What did he get?" The father said, "Lou Gehrig disease." "Wow!", the son responded, "What are the odds?!!"

[Slide 6] There has been considerable discussion of the causal mechanisms behind the association of health and SES, but relatively few natural experiments that permit causal paths to be definitively identified. Today, I will discuss research that advances the investigation by testing for the *absence* of direct causal links. My sources are results obtained by my team, published in the Journal of Econometrics last January, and in a forthcoming addendum, and three related papers, by Michael Marmot and co-authors, by Anne Case and Angus Deaton, and by Jim Smith.

[Slide 7] This figure depicts possible causal paths for the health and SES innovations that occur over a short period. An individual's life history is built from these period-by-period transitions. First, low SES may lead to failures to seek medical care and delay in detection of conditions, reduced access to medical services, less effective therapies, or failures to maintain treatment regimens. Also, increased risk of health problems may result from increased stress or frustration, or increased exposure to environmental hazards, that are associated with low SES. These factors could provide direct causal links from SES history to health events. Second, poor health may reduce the ability to work or look after oneself, and increase medical expenditures, leading to reduced income and less opportunity to accumulate assets. This could provide a direct causal link from health to changes in SES.

There may also be hidden *common factors* that lead to ecological association of health and SES. For example, unobserved genetic heterogeneity may influence both resistance to disease and ability to work. Causal links may be reinforced or confounded by behavioral response. For example, rational economic decision-making may induce robust consumers to accumulate in order to finance consumption over a long expected retirement, or unhealthy individuals with diminished horizons to spend down assets. Behavioral factors such as childhood nutrition and stress, exercise, and smoking may influence both health and economic activity level. I recently described some of the links between health and wealth to Paul Samuelson. He said that it is obvious, different social classes think differently about the future. He told me the following story. In World War II, the economist Frank Holtzman was an officer on a ship approaching the Suez Canal on its way to Buma. The captain called assembly before the ship docked in Alexandria, and warned the crew that syphilis was so rampant in Alexandria that your nose might fall off before the ship left port. The captain advised that the port of call beyond the canal was much healthier. Holtzman reported that when the ship docked in Alexandria, all the officers cowered on board, while all the enlisted men headed for shore, saying it's no worse than a bad cold. Another story also makes the point that failure to plan ahead can lead to bad health outcomes. A poor man delays going to his doctor, and when he finally does, the doctor returns with the lab results, and says: "I have very bad news, you have ten." The man exclaims: "My God, Doc, what? Ten years, ten months?" The doctor says: "Ten..nine..eight..."

[Slide 8] This rather busy slide details the causal paths one might expect to find. Health events can be broken down into onset, diagnosis, choice of treatment, and outcomes, which may be a cure, control of a chronic or degenerative condition, or death. Common factors, here labeled "behavior", could influence the onset of some health conditions, such as type II diabetes, and could influence how conditions are diagnosed and treated, and compliance with treatment regimens and the consequent effectiveness of treatment.

SES will have a direct causal link to diagnosis and treatment if the health care delivery system favors those with social or financial resources. These arrows are marked in red and will be of particular interest in our analysis. Finally, health conditions and their treatment entail out-of-pocket medical and drug costs, cost of caregiver services, and disability-induced income reductions or job loss. In countries and demographic groups with comprehensive health insurance, direct out-of-pocket medical costs may be limited, but few systems cover all the costs of living arrangements that health conditions may necessitate.

[Slide 9] Much of the interest in the health/SES nexus is fueled by the potential policy implications of alternative causal paths. If the poor are less healthy because of specific failures in the health care delivery system, targeted changes in that system could have a significant net social benefit.

If the poor are less healthy because of behavior that increases risk, delays diagnosis, or reduces the effectiveness of treatment, then policies that increase information and encourage protective behavior, or increase preventative interventions, may be beneficial. There may be an important distinction between direct causal mechanisms influencing mortality, conditioned on health status, and direct causal mechanisms influencing onset of health conditions. For mortality, an SES gradient could be due to differentially effective treatment of acute health conditions. For morbidity, an SES gradient could reflect differentials in prevention and detection of health conditions. These involve different parts of the health care delivery system, and differ substantially in the importance of individual awareness and discretion, and allocation of costs between health insurance and the individual. Finally, if the less healthy are poor due to work disability and medical costs, then health and disability insurance need to be strengthened.

[Slide 10] The research we have done on deconstructing the health-wealth nexus and identifying direct causal links has focused on a panel of elderly Americans. In this sample, we test for the *absence* of direct causal paths from SES to innovations in health, and from health status to innovations in SES. This path diagram, also called a directed acyclic graph, illustrates possible paths. Health states, denoted H , and socioeconomic states, denoted S , evolve through time, with the arrows denoting the direct causal paths. Our null hypotheses are that there are no direct *causal* paths from S_{t-1} to H_t , or from H_{t-1} to S_t ; i.e., the red causal arrows are absent. What we are actually able to test is the *joint* hypothesis that there is no direct causal path from S_{t-1} to H_t *and* no hidden common factors that could induce an ecological correlation of S_{t-1} and H_t even in the absence of a direct causal path.

[Slide 11] Here is the same path diagram under our null hypothesis in which a direct causal link from S_{t-1} to H_t is absent. Note that the absence of a direct causal path does not rule out *indirect* causation, in which an association of S_{t-1} and H_t is transmitted from H_{t-2} to S_{t-1} and, through H_{t-1} , to H_t . However, conditioning H_t on the observed variable H_{t-1} accounts fully for this indirect causation, so that *conditioned on* H_{t-1} , no association of S_{t-1} and H_t will be found unless a direct causal path is present.

[Slide 12] Here is another alternative to the compound null hypothesis we actually test, in which there is a persistent unobserved common factor C . We follow the convention from path analysis that this unobserved factor is denoted by an open node, while observed variables are denoted by filled nodes. Indirect association between variables is blocked by a filled node, as conditioning on the observed variable at this node captures all this association. However, correlation passes through open nodes. These useful syllogisms for path diagrams are nicely summarized in Judea Pearl's book on causality.

[Slide 13] In our case, even in the absence of a direct causal link from S_{t-1} to H_t , correlation passes from S_{t-1} through C_{t-1} and C_t to H_t , so that S_{t-1} is predictive for H_t despite conditioning on H_{t-1} . Control function methods using instruments such as changes in tax rates or Medicaid coverage are needed to identify direct causal effects in the presence of common factors; this is a topic of ongoing research.

[Slide 14] Let me now describe how we do our analysis, and some of the methodological considerations that go into it. The health and socioeconomic states that we work with can be treated as discrete, so that the evolution of health, wealth, and other variables can be modeled as a discrete-state stochastic process. We can approximate this as a first-order Markov process, provided we are sufficiently generous in describing states, and allow for the presence of hidden or unobserved states. A complete causal analysis would proceed by specifying a full *structural model* in the language of econometrics, or a full *potential outcomes analysis* in the language of statistics. This is a powerful approach, giving invariance properties that allow valid predictions of the effects of broad ranges of policy interventions. However, it is hard to get a fully specified model right, and hard to verify its validity.

[Slide 15] What we do instead is set a more limited objective, to specify Markov models that meet invariance conditions that are just sufficient for the validity of targeted policy analyses. A model is *structural* relative to a specified family of histories if it has the invariance property that it is valid for each history in the family. Operationally, invariance means that within specified history and treatment domains, f has the *transferability* property that it is valid in different populations where the marginal distribution of H_{t-1}, S_{t-1} changes, and the *predictability* or *invariance under treatments* property that it remains valid following policy interventions that alter the marginal distribution of H_{t-1}, S_{t-1} .

[Slide 16] Restricted history and treatment domains identify families of structural models, or equivalently families of path diagrams. I would term this a *bottom up* approach to causal analysis, in which one starts from a relatively narrow family of invariance tests that are empirically practical and policy-relevant, and makes limited claims on the features of the family of consistent structural models, in contrast to a *top down* approach in which one specifies a full causal model which can, however, be only partially validated with available evidence. In econometric terms, our approach is analogous to specifying a single equation in a simultaneous equation system, and only reduced form equations for the remaining endogenous variables, but here we look for invariant or structural features that hold for a whole vector of endogenous variables

within a limited domain of histories, rather than a structural feature that holds for a single endogenous variable under an unlimited domain of histories. Within our framework, the steps for testing for the absence of direct causal paths are straightforward. We first test for invariance, and when invariance is accepted, so that a case is made that the model is valid for the relevant policy interventions, we do simple exclusion tests on the Markov transition probabilities to determine if SES contributes to the predictability of health innovations, or health state contributes to the predictability of wealth innovations. These can be interpreted as Granger tests for non-causality, but together with the invariance tests they are potentially much more useful than simple Granger tests in time series models that may not be structurally stable. We believe our approach is a scientifically useful way of thinking about causality and causality testing. However, it has some serious limitations. Our hypotheses will in general be accepted *only* if no causal link is present *and* there are no persistent common factors that influence both initial state and innovations. Rejection of one of these hypotheses does not demonstrate a direct causal link, since this may be the result of common hidden factors. In addition, our tests have no power against the alternative of a direct causal link that is present but is effectively masked by offsetting common factors within the domain of histories we consider. Using genetic language, we cannot detect direct causal paths that have not been expressed, and are not active in the domain that we analyze. There is an analogy here to exogeneity testing in econometrics, often a useful exercise even though positive power against every alternative is not usually guaranteed.

[Slide 17] In our analysis, we work with 22 health conditions and five time-varying socioeconomic variables. We array these in a Wold causal chain suggested by the etiology of the health conditions. In this model framework, we test for model invariance across waves of the panel, and for the causal chain structure. This is considerably short of the battery of invariance tests that would be needed for useful policy analysis, and for that reason some of the work of Marmot and Smith that I will discuss later is particularly useful. Finally, with invariance as a maintained hypothesis, we conduct the exclusion or Granger tests for non-causality.

[Slide 18] I will turn now to the data that we use in our study, and our empirical results. We use the AHEAD panel, now part of the Health and Retirement Study. This panel consists of 7447 individuals aged 70 and above in 1993, including spouses. The panel is generally representative of the population, with oversampling of non-whites and Floridians; there is nevertheless insufficient data to estimate the same models for non-whites that we do for whites. We use the first three waves, from '93, '95, and '98. Recently, a fourth wave was released, but is not included in the results I will discuss today. An initial question one might ask is whether this is a good population in which to look for direct causal paths. This population is retired, so that health events should have little impact on earnings. Most are on Medicare, which limits out-of-pocket costs for most medical treatments, except for drugs. Then, direct links from health conditions to SES should operate primarily through the influence of drug and caregiver costs on net savings rates. As my first table of relative risks showed, most of the association of health and SES is already built into initial state when you reach age 70. Consequently, it would not be surprising if direct causal links in this population were weak or absent. Statements about the presence or absence of direct causal mechanisms in this

population, given previous health and SES status, say nothing about the structure of these mechanisms in a younger population, where associations of health and SES emerge as a result of some pattern of causation and operation of common factors. One might also ask whether common factors are likely to be so common in the AHEAD population that the no-causation hypotheses will almost always be rejected, leaving the issue of direct causation unresolved. Here there is some cause for optimism. In an elderly population, persistent hidden factors will often be expressed through observed covariates, so that once these covariates are controlled, the residual impact of the hidden factors on innovations will be small. For example, genetic frailty that is causal to both health problems and low wages, leading to low wealth, may be expressed through a health condition such as diabetes. Then, onset of new health conditions that are also linked to genetic frailty may be only weakly associated with low wealth, once diabetic condition has been entered as a covariate. Thus, in this population, *rejection* of the hypotheses *may* provide useful diagnostics for likely causal paths.

[Slide 19] I will turn now to our results. This slide and the next summarize our findings for the critical invariance and non-causality tests. The invariance that is being tested here is parameter invariance across waves. In this slide, we examine major health events, such as cancer, a heart attack, stroke, death, lung disease, and diabetes. We find for these major events, the models pass our invariance test, and we accept the hypothesis of non-causality of SES. These results support the view that given initial health status, SES is not protective against the onset of major health conditions, and there is no SES gradient in the availability or effectiveness of treatment that induces a gradient in mortality.

[Slide 20] For the chronic diseases, accidents, and mental conditions in this slide, the pattern is a little more ambiguous. Onsets of cognitive impairment, psychiatric disease, depression, and poor or fair SRHS show an SES gradient in most cases. As I indicated earlier, this may be due to common factors, so one cannot conclude that a direct causal link is responsible. However, it is suggestive that the evidence for a causal link is in the area of mental illness where drug and caregiver cost coverage in Medicare is most restricted, and where individual behavior is important for diagnosis and for the effectiveness of treatment.

[Slide 21] Another way to view our results on direct causal links is to look at the relative risks of low versus high SES for various health conditions. As in a previous table for prevalence relative risks, we are defining low SES to be bottom quartile in income and wealth, less than a high school education, and a poor dwelling and neighborhood, and high SES to be top quartile in income and wealth, a college education, and a good dwelling and neighborhood. Again, the relative risks are calculated holding concurrent health conditions, demographics, and smoking history constant. What is significant now is the relatively high risk of low SES for lung disease for females, and mental illnesses and SRHS for both males and females.

[Slide 22] Summarizing our results, we find that cross-wave invariance holds for most conditions. We test the causal chain structure that we have imposed. Generally, it is supported, but there are some notable exceptions, particularly common unobserved

factors entering ADL's, IADL's, accidents, cognitive impairment, and mortality. We find that for acute conditions and mortality, the hypothesis of no direct causal links from SES is accepted. For some mental conditions and SRHS, the hypothesis of no direct causality is rejected, suggesting that either behavioral common factors or direct causal links are operating.

[Slide 23] Our analysis of SES to health pathways in the age 70+ population provides no evidence on links in younger working-age populations. Case and Deaton find that through the working years, health as measured by SRHS deteriorates with age more rapidly for manual laborers than for other occupations. There are some measurement issues associated with SRHS; nevertheless, these results suggest that exposure to manual labor is a risk factor for health.

[Slide 24] Our tests for direct causal links from health status to changes in wealth are inconclusive. Generally, our models do not satisfy cross-wave invariance. This may be substantially due to undercounting of assets in AHEAD Wave 1, and to measurement errors in the wealth variables. Our non-causality test fails for liquid wealth, and in some cases for non-liquid wealth, but absent invariance it is difficult to say whether direct causal links or common factors are at work.

[Slide 25] The generality of our results is limited by our concentration on an elderly population, and execution of a single test for invariance across waves. This limits our ability to make inferences about the impact of policy interventions, such as addition of drug benefits to Medicare, for which invariance has not been tested, or to make inferences about direct causal paths in younger populations. For these reasons, a study by Adda, Chandola, and Marmot is particularly revealing. These authors replicate our tests for non-causality of SES for incidence of new health conditions on Whitehall II data and on Swedish ULF data, which contain younger cohorts than AHEAD. They obtain essentially the same results that we get in the AHEAD data. We interpret these results as supporting, broadly, a transferability invariance property across quite different health care delivery systems as well as younger cohorts. The slide shows significance levels for non-causality tests for selected health conditions. In addition to providing comfort that our models are reasonably specified, these results suggest strongly that the primary source of the failure of the non-causality test for mental conditions is behavioral or genetic, rather than a direct causal link from SES.

[Slide 26] As we did for prevalence and incidence of health conditions in AHEAD, we can calculate the relative risk of low versus high SES from the Whitehall and ULF studies. First, note that the definitions of SES states are country-specific, both conceptually and in terms of survey measurements. This is presumably the reason that the ULF relative odds are more extreme than those in AHEAD. In addition, there is more variability in relative risk across the three studies than there was in non-causality test significance levels. Nevertheless, a broad pattern emerges that relative risks are highest for SRHS and chronic conditions, and lowest for acute illnesses. The table does not show mental conditions, but they also have high relative risks in both AHEAD and ULF, and moderately high in Whitehall. More interesting work could be done across these three

panels, but the initial results provide encouraging support for our methodology, and the potential for our models to satisfy invariance tests across interesting policy dimensions.

[Slide 27] One of the disappointing aspects of our study, particularly from an economist's point of view, is that it failed to give conclusive results for tests for direct causal links from health status to innovations in wealth. In a recent paper, Jim Smith has studied younger cohorts in the Health and Retirement Survey, aged 51-61 in 1992, and tracked carefully the cumulative income loss and out-of-pocket expenses associated with a major or minor health event occurring between the first two waves of the study. This slide gives selected quantiles for net out-of-pocket medical costs. A first notable observation is that while out-of-pocket costs are modest on average, compared to income or wealth, the upper tail for these costs is a substantial fraction of the total assets of a typical household. For example, the 95th percentile of net cumulative OOP medical expenditures associated with a major health event is about one-fourth of the total assets of a typical household. Smith has looked at whether health insurance matters for out-of-pocket costs. He finds that average out-of-pocket outlays are comparable at the mean, but that those without health insurance have a higher upper tail for large outlays.

[Slide 28] Lost income as the result of a major health problem looms even larger than out-of-pocket costs as a direct impact on the ability of individuals to accumulate assets. Since the years from age 50 to the early 60's are a prime period for retirement accumulation, these findings indicate that we can expect to see significantly lower wealth for AHEAD aged households when a member has had a major health problem during his or her 50's. Smith finds that major health problems are much more likely to lead to job loss at low income levels than at high income levels, which amplifies the SES gradient. Of course, one cannot assume the combined lost income and medical expenditures associated with a major health event come entirely from savings. Nevertheless, at historical market rates of return for assets, Smith's findings imply that a substantial part of the wealth gradient for prevalence of major health conditions observed among 70 to 75 year-olds in AHEAD can be attributed to the direct impact of major illnesses on savings behavior ten to twenty years earlier. This suggests that however important and policy-relevant early life circumstances and nutrition may be for health late in life, their effect may be amplified by a more proximate and direct source for much of the observed SES gradient in prevalence rates, a direct link from health to out-of-pocket expenses and lost earnings in the final decades of working life when health hazards rise and opportunities for retirement savings are the greatest.

[Slide 29] One feature of the Markov models for health and SES that we have estimated is they can be used to simulate the life histories of a synthetic population, and predict the impact of policy interventions that influence the incidence or outcomes from specific diseases. Such simulations can account consistently for competing and complementary hazards, and predict the overall impact of policy interventions on survival. For analysis of health policy, a holistic picture of live histories is a useful complement to the partial equilibrium disease-centric focus of most medical and epidemiological studies. Of course, to do this reliably requires that we estimate consistently all the direct causal links. This requires estimation methods beyond what we have done so far, with the use

of instruments and control functions to overcome the ecological correlations introduced by common factors. However, for illustrative purposes, we have constructed life history simulations with our existing models. Here, for example, are results for survival probabilities for females and males, under baseline conditions and following a hypothetical medical intervention that eliminates diabetes in our starting cohort of 70 year olds, although not the other health conditions they may have accumulated at this age as a result of diabetes in the past. According to this simulation, life expectancy from age 70 would be increased by 5 months for females and 7 months for males if diabetes were eliminated.

[Slide 30] In the U.S. retired, Medicare-eligible population, the evidence is against strong direct causal links from SES to incidence of most new health conditions, or to mortality. Self-reported health status and mental conditions show an SES gradient, with the cross-country evidence favoring individual behavior as the source, rather than deficiencies in delivery of mental health services.

[Slide 31] There is inconclusive evidence for direct causal links from health to wealth changes within the AHEAD panel, but occupational exposure, income loss and medical costs suggest strong links in working-age populations. Even though relative risks are mostly near one, SES gradients operating from age 70 to the end of life have an economically significant cumulative effect on incidence and mortality. In conclusion, we have focused a little the search for the direct causal paths that through the life course induce the association of health and SES, and provided some methodological tools that we hope will facilitate further work.