HEALTH DISPARITIES IN MID-TO-LATE LIFE: THE ROLE OF EARLIER LIFE FAMILY AND NEIGHBORHOOD SOCIOECONOMIC CONDITIONS*

RUCKER C. JOHNSON
GOLDMAN SCHOOL OF PUBLIC POLICY
UNIVERSITY OF CALIFORNIA, BERKELEY

ROBERT F. SCHOENI
INSTITUTE FOR SOCIAL RESEARCH &
FORD SCHOOL OF PUBLIC POLICY
UNIVERSITY OF MICHIGAN

JEANNETTE ROGOWSKI
SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MEDICINE & DENTISTRY OF NEW JERSEY

MAY 2009

* Please direct correspondence to Rucker Johnson, University of California, Berkeley, Goldman School of Public Policy, 2607 Hearst Ave, Berkeley, CA 94547, or email to ruckerj@berkeley.edu. This research was supported by funding from NIA grant R01AG04058-02. We wish to thank the PSID staff for access to the confidential restricted-use PSID geocode data.

HEALTH DISPARITIES IN MID-TO-LATE LIFE: THE ROLE OF EARLIER LIFE FAMILY AND NEIGHBORHOOD SOCIOECONOMIC CONDITIONS

Abstract:

This paper investigates the degree to which childhood SES and young adult family, neighborhood, and individual-level factors contribute to adult health disparities in mid-to-late life. We use correlations based on a nationally representative longitudinal sample of married couples and neighbors followed from young adulthood through elderly ages to estimate bounds on the possible causal effects of young adult family and neighborhood characteristics on general health status in mid-to-late life. Estimates based on four-level hierarchical random effects models consistently show a significant scope for both young adult family and neighborhood factors. The estimates suggest that disparities in neighborhood conditions experienced in young adulthood account for up to one-quarter of the variation in health status in mid-to-late life. While the neighbor correlations must be strictly interpreted as upper bounds, the estimates suggest that neighborhood factors experienced earlier in the life course influence both contemporaneous and future health outcomes. In particular, we find that living in a high poverty neighborhood during young adulthood has large harmful consequences on mid-to-late life health, while contemporaneous neighborhood poverty is only weakly related to mid-to-late life health. To probe the robustness of a causal inference, the analysis employs a novel empirical approach, recently proposed by Altonji et al. (2005), to gauge how sensitive estimates of the effects of neighborhood poverty are to selection on unobserved variables. The results reveal that even a large amount of selection on unobservable factors does not eliminate the significant effect of neighborhood poverty on health status later in life. Finally, the results indicate that childhood SES and young adult neighborhood and family factors can account for three-quarters of the black-white gap in health status at ages over 55.

I. INTRODUCTION

Increasingly, research that focuses on health in aging populations is recognizing the need to incorporate earlier life circumstances in analyses of adult health. At the same time, unraveling the sources of racial and socioeconomic differences in adult health outcomes has eluded studies that have attempted to explain them using only individual-level factors and contemporaneous socioeconomic factors. The dearth of available longitudinal, nationally-representative US data with extensive information on socioeconomic status and health status has presented formidable challenges to understanding how early, middle, and late life factors influence the life cycle trajectory of health. This paper analyzes the influence of lifecycle socioeconomic achievement on the health experiences of individuals in mid-to-late life, and it builds on prior work to contribute to our understanding of how health and socioeconomic attainment processes are linked across childhood, young adulthood, middle-age, and late-life.

We investigate the extent to which later-life health inequality is related to the different types of contexts into which individuals are born, and within which they grow up, live, and work at earlier stages of the life cycle. Understanding the magnitude, nature, and multiple dimensions of the socioeconomic gradient in health and sources of racial health differences over the life course may be wholly missed in analyses that are confined only to traditional individual and family-level variables and contemporaneous economic measures, without consideration of the influence of earlier life neighborhood environments and community context. Accordingly, this study uses data from the Panel Study of Income Dynamics (PSID), spanning nearly four decades, to investigate whether and how neighborhoods affect the aging process over the life course.

There is strong theoretical rationale supporting the hypothesis that neighborhood factors influence health, and several empirical studies have found strong associations between health and specific neighborhood variables (e.g, Ellen et al., 2001; Morenoff and Lynch, 2002; Chandra and Skinner, 2003). The typical analytical approach used in these empirical studies is to regress individual level outcomes such as education, criminal activity, or health on contemporaneous neighborhood-level factors such as census tract mean income, poverty rates, or rates of single motherhood. But attempts to estimate causal effects of neighborhood context have faced well-documented challenges of endogeneity (Manski, 1993, Oakes??) and of obtaining accurate measures of neighborhood factors. Few studies have used convincing identification strategies to overcome these challenges, exceptions being experimental evaluations such as Katz, Kling, Liebman (2001) and Leventhal and Brooks-Gunn (2001).

This paper employs a different approach to address these challenges by exploiting unique features of the PSID. Specifically, the first objective of this study is to bound the proportion of inequality in late-life health that may be attributed to child socioeconomic status and disparities in neighborhood and family characteristics during early-to-mid adulthood. The research strategy exploits the fact that the initial PSID sample in 1968 was highly clustered, with most PSID families having several other sample families living on the same block. This survey design allows us to compare the similarity in mid-to-late-life health between spouses, versus unrelated individuals who were living in the same narrowly defined neighborhood in young adulthood. This approach avoids the difficulty of defining neighborhood quality at the outset, and instead compares the resemblance in health status in mid-to-late life between spouses and adult neighbors, placing an upper bound on the neighborhood influence. The comparison of spousal correlations with adult neighbor correlations in mid-to-late life health allows an assessment of the relative magnitudes of the effects of the neighborhood environment in adulthood versus family characteristics in adulthood. Spousal correlations in health reflect the influences of shared household resources and neighborhood environments. Relative to the correlation among spouses, do adult neighbors have highly correlated health status? Small adult neighbor correlations would indicate that adult neighborhood factors can explain only a minor portion of the variation in later-life health outcomes. Large neighbor correlations would leave open the possibility that neighborhoods contribute significantly to inequality in health outcomes, and further analyses of the effects of particular neighborhood characteristics would be warranted.

The analysis is carried out in three stages. First, we estimate how much childhood SES and young adult family, neighborhood, and individual-level factors (observed and unobserved) contribute to adult health disparities in mid-to-late life. The results are based on the estimation of four-level hierarchical random effects models of health status and (as aforementioned) the analysis involves the comparison of spousal correlations in mid-to-late life health with young adult neighbor correlations in mid-to-late life health (net of the similarity arising from neighbors having similar family characteristics).

Second, after providing evidence highlighting a significant overall scope of neighborhood and family socioeconomic factors earlier in life on subsequent health trajectories, we attempt to identify the specific childhood and young adulthood family and neighborhood-level characteristics that influence later-life health. We focus on socioeconomic family and neighborhood conditions experienced at least twenty years prior to the health status outcomes that we examine using an array of observable neighborhood and family factors measured earlier in the life course. We consider the duration of exposure to disadvantaged/advantaged neighborhood environments over the life course, and the relationship between cumulative neighborhood exposures over the life course and later-life health. The results demonstrate what aspects and sources of later-life health disparities may be missed using traditional models that focus more on contemporaneous socioeconomic factors, without the emphasis on earlier life factors.

Third, after uncovering evidence that concentrated neighborhood poverty is associated with significantly elevated risks of problematic health, even after controlling for an extensive set of individual and family-level factors, we conduct a sensitivity analysis to probe the robustness of the results for causal inference. Specifically, we use a novel empirical approach, recently proposed by Altonji et al. (2005) and Krauth (2007), to test the robustness of the estimated effects of young adult neighborhood poverty to selection bias due to an omitted variable. The goal is to assess how the point estimate and confidence interval of the effect of neighborhood poverty change under the presence of selection bias of varying strengths. This analysis allows one to determine the threshold of selection on unobservables, if any, at which neighborhood poverty during young adulthood no longer has a significant effect on adult health. Finally, our empirical analysis makes a further unique contribution in producing evidence on the role of neighborhood environments in contributing to socioeconomic and racial health disparities later in life.

The remainder of the paper is organized in the following way. We begin with a discussion of the life course perspective of health that incorporates the ways in which neighborhood and family background may matter. This section provides the conceptual framework, highlights the relevant theoretical issues, and motivates the empirical analyses to follow. Section III lays out the methodological challenges in estimating neighborhood effects. The empirical approach is outlined in section IV and the data are described in section V. Section VI discusses the econometric model and estimation methods. The results are presented in section VII, with concluding statements provided in the final section.

II. CONCEPTUAL FRAMEWORK AND THEORETICAL CONSIDERATIONS

This paper takes a life-course perspective of health and hypothesizes that inequalities in health status are a product of the interaction between the developmental opportunities and vulnerabilities at each stage of the life course, on the one hand, with widely varying attributes of socioeconomic neighborhood and family conditions, on the other. The life-course approach applied to health posits that the long-term health consequences of earlier life socioeconomic adverse conditions may manifest through three possible processes: “latency”, “cumulative”, and “pathway” processes. *Latency* refers to resultant effects of adverse conditions early in life that cause the body’s physiology and metabolism to be fundamentally altered, with some of the consequences of these changes manifesting in visible health problems much later in life. The fetal origins hypothesis is the most prominent among the set of potential latency mechanisms (See Barker, 1998, for a review; and Johnson & Schoeni, 2007, for supporting evidence). Recent findings in neuroscience indicate that developmental health trajectories can be altered more readily in childhood during sensitive periods of rapid developmental change than during other periods. Investments in early life influence the accumulation of health capacity during childhood and the early developmental experiences build a foundation that may influence the return to subsequent human capital investments and the later-life rate of decline. Childhood poverty exposures may, thus, lead to a damaged health stock making it more susceptible to deterioration later in life. Alternatively, health shocks early in life, even in the womb, may have immediately visible health effects that are chronic, lasting from birth through adulthood and old-age.

The hypothesis that the differential burden of lifetime stress contributes to racial and socioeconomic disparities in health is closely related with the *cumulative* process explanation.

Persistent exposures to disadvantaged neighborhood and family conditions may have a cumulative toll in the form of "weathering" (Geronimus, 1992). This conceptual framework is consistent with a stress and adaptation perspective on how stressful neighborhood conditions may influence health trajectories. Prolonged exposure to stress produces elevated risks of a condition known as allostatic load, which refers to the physiological costs of chronic overactivity or underactivity of systems within the body (e.g., the hypothalamic-pituitary-adrenal axis or the autonomic nervous system) that fluctuate to meet demands of chronic exposure to environmental stressors (McEwen, 1998). How people are affected and adapt to stressful neighborhood environments may depend, in part, on their access to informal sources of social support.

Pathway processes refer to linkages between health and socioeconomic attainment processes where childhood socioeconomic neighborhood and family factors affect adult socioeconomic outcomes such as education which in turn influences health later in life (Kuh and Wadsworth, 1993). The degree of socioeconomic mobility has direct implications on the resemblance of an individual's childhood, young adulthood and mid-to-late adulthood family characteristics, such as income and education, which may in turn affect health. Since economic status is a major determinant of residential choice, persistence in economic status is likely to lead to persistence in neighborhood quality as well; that is, the lower economic mobility is, the greater the correlation between childhood and young- and mid-to-late adulthood neighborhood characteristics.

Latent, cumulative, and pathway explanations for the linkages between health and socioeconomic status are not mutually exclusive and may produce synergistic influences. While separately identifying how latent, cumulative, and pathway effects combine with current circumstances to impact later-life health is beyond the scope of this paper, the underlying framework informs the modeling approach taken in our empirical work.

III. METHODOLOGICAL CHALLENGES IN ESTIMATING NEIGHBORHOOD EFFECTS

The key methodological hurdle that studies in the neighborhood effects literature must confront is uncovering the source and nature of the association between neighborhood factors and the outcome of interest, in our case health status. Unobserved factors that affect health may also be correlated with neighborhood factors, leading to biased estimates of neighborhood effects. This can arise from the endogeneity of residential location. That is, individuals and families choose where they live based on the characteristics they value (Tiebout, 1956), although constraints such as racial discrimination and exclusionary zoning may be placed on that decision. In this context, families and individuals who care more about their health will be less likely to choose to live in an area with high crime, pollution, or a poor health care system. Furthermore, the set of complex and nuanced characteristics that influence neighborhood choices are not likely to be well measured and accounted for appropriately in econometric models.

Given these challenges, social scientists have yet to make much ground on producing more convincing evidence on the impact of neighborhoods on individual outcomes. Moreover, the typical methods used by microeconometricians to address endogeneity problems (e.g., instrumental variables and fixed effect approaches) have significant limitations in this context. First, most health outcomes are a product of cumulative exposures to advantaged/disadvantaged environments spanning decades or exhibit long latent periods before problems manifest. Therefore, the connection between current neighborhood and current health may say little about the overall influence of neighborhood factors over the life cycle. As well, it may be important to conceptualize neighborhood effects as cumulative and variable over the life course as opposed to isolated and unchanging. Because most methods for overcoming endogenous neighborhood choice are based on small, short-run changes in the neighborhood environment, these approaches might be limited to uncovering effects only for rapidly-responding intermediate outcomes such as health behaviors (e.g., smoking/drinking, exercise/diet). An additional issue is that neighborhood variables of the underlying neighborhood feature of interest are notoriously measured with a great deal of noise. The neighborhood attributes of interest change slowly over time, so most year-to-year variations in the characteristic measured are noise.

The most powerful way to address selection is through a randomized trial. But an experimental design where neighborhoods are randomly assigned is rare. A significant exception is the evaluation of the Move to Opportunity (MTO) program, where an experimental design is used to estimate the effects of offering housing assistance that allows individuals to move out of low-income, poor neighborhoods. Evidence from two sites – Boston and New York – demonstrates that MTO had beneficial effects on the health of children and adults – specifically, adults in the experimental group (those who were given vouchers to move to more affluent neighborhoods) experienced improvements in self-rated health and mental health after they moved, and children in these families experienced fewer injuries and a lower probability of having an asthma attack (Katz, Kling, Liebman, 2002; Leventhal and Brooks-Gunn, 2002). This evidence is consistent with the claim that neighborhood factors do in fact influence health status, at least in the short-run among poor families.

Among the studies that have tried to address endogeneity and self-selection using non-experimental methods, the most common approach is the use of instrumental variable techniques (e.g., Evans et al., 1992; Case and Katz, 1991; and McLanahan, 1996), where the exclusion restrictions are tenuous. An alternative non-experimental approach is comparing siblings who have been raised in different neighborhoods at different ages because their parents have moved (Aaronson, 1998; Plotnick and Hoffman, 1996). The key assumption is that the family effect is fixed, not time-varying. If, for example, families' preferences change as their children get older, and they become more interested in living in neighborhoods that are less risky for their children's health, then they might move to neighborhoods with less crime or pollution, which may in turn lead to better health outcomes for their kids. However, if the underlying change in their preferences towards health outcomes caused them to not only change neighborhoods, but also spend more time encouraging their children to practice good health behaviors such as eating healthily, exercising, and avoiding high crime areas, then the neighborhood "effect" might actually be representing all of these other factors and not the true causal effects of neighborhoods *per se*. Moreover, it is quite possible that sibling differences may aggravate the endogeneity problem, as has been discussed in the context of the labor market returns to schooling (Griliches, 1979; Bound and Solon, 1999).

Typical neighborhood studies also face the challenge of identifying and measuring relevant factors. The neighborhood qualities that may in fact matter may be hard to measure, or they may not be measured in enough spatial detail (see, e.g., Kawachi and Berkman's (2003) discussion of measurement issues of neighborhood attributes). This issue is analogous to the finding in the family background literature that sibling correlations in socioeconomic status far exceed what has been explained by any particular measured aspects of the siblings' shared background (Corcoran, Jencks, and Olneck, 1976).

IV. EMPIRICAL APPROACH

Instead of performing another regression analysis focused on particular neighborhood characteristics, in this paper we exploit a unique feature of the PSID and modify an approach used by Solon et al (2000) to examine the role of childhood neighborhood factors on educational attainment, and more recently used by Johnson (2008) to examine the role of childhood contextual factors on health through mid-life. Specifically, the initial PSID sample in 1968 was highly clustered with most PSID families having several other sample families living on the same block, who have been subsequently followed over time. We follow the health experiences of those who were in their 20s and 30s in 1968, and thus who had reached (or were approaching) elderly ages by 2005. For our purposes, this survey design allows us to compare the similarity in mid-to-late-life health between spouses, versus unrelated individuals who were living in the same narrowly defined neighborhood during young adulthood.

The basic idea is to estimate the correlation of health status of adults who were neighbors earlier in their life. We assess the extent to which health status is correlated among neighbors above and beyond the correlation that arises due to family effects (i.e., net of the similarity/resemblance because neighbors have similar family characteristics due to residential sorting). Because all 1968 family members within a given family are followed throughout their lives, the correlation among neighbors can be adjusted to account for family-specific factors. This approach avoids the difficulty of defining neighborhood quality at the outset, and instead asks: relative to the correlation among spouses, do adult neighbors have highly correlated health status? The comparison of spousal correlations with young adult neighbor correlations in mid-to-late life health allows an assessment of the relative magnitudes of the effects of the neighborhood environment in adulthood versus family characteristics in adulthood, placing an upper bound on the neighborhood influence.

Unlike Solon et al (2000), this paper's research design uses spouses instead of siblings to identify the family component. Siblings cannot be used to study the family component in later-life health because the PSID sample members who were children when the PSID began in 1968 have not yet reached old age. However, young adults and their spouses as of 1968 have attained older ages. Spousal correlations in health reflect the influences of shared household resources and neighborhood environments. Small adult neighbor correlations would indicate that adult neighborhood factors can explain only a minor portion of the variation in later-life health outcomes. Large neighbor correlations would leave open the possibility that neighborhoods contribute significantly to inequality in health outcomes, and further analyses of the effects of particular neighborhood characteristics would be warranted. The results are based on the estimation of four-level hierarchical random effects models of health status.

There are four primary reasons why the approach taken in this paper may be able to detect neighborhood effects in ways previous studies have been unable to do. First, in contrast to the experimental evidence and previous observational studies, the analysis examines effects over a much longer time horizon. This is particularly important for most health outcomes, as there is likely a substantial lag between poor neighborhood quality and the manifestation of health effects. Second, instead of focusing on contemporaneous neighborhood effects, we are analyzing the effects of neighborhood characteristics in young adulthood on health in older ages, which will include indirect effects operating via the economic mobility process as well as cumulative exposure to neighborhood conditions that may vary over the life cycle. Causality between economic and health status runs in both directions in adulthood and has proven notoriously difficult to sort out empirically (Adda et al., 2003; Smith, 1999). By focusing on socioeconomic neighborhood conditions experienced at least twenty years prior to the health status outcomes that we examine, we can more easily identify causal impacts on subsequent health because reverse causality is a lesser concern. Third, we use the census block as the definition of neighborhood, which comprises a much smaller geographic area than previous studies utilize. Finally, we use estimates of neighbor correlations as an omnibus measure of the potential effects of neighborhood quality (including unmeasured characteristics), rather than initially focusing the analysis on particular observable neighborhood attributes.

In addition to providing evidence on the overall scope of neighborhood and family socioeconomic factors earlier in life on subsequent health trajectories, our empirical analysis makes two further unique contributions on: (1) the relationship between cumulative neighborhood exposures over the life course and later-life health; and (2) the role of neighborhood environments in contributing to socioeconomic and racial health disparities. Our innovative research design and unique measures collected on aspects of neighborhood physical, service and social environments—including neighborhood poverty and crime, income and education, health insurance, race and residential segregation, health behaviors, housing quality, connectedness to informal sources of support—help illuminate what lies along the "chain of causation" from poverty to later-life health outcomes.

V. DATA AND MEASURES

The PSID began interviewing a national probability sample of families in 1968. These families were re-interviewed each year through 1997, when interviewing became biennial. All individuals are followed until death or attrition, and data through the 2005 wave is utilized in the analysis. The PSID maintains extremely high wave-to-wave response rates of 95-98%. Studies have concluded that the PSID sample of heads and wives remains representative of the national sample of adults (Gottschalk et al, 1999; Becketti et al, 1997).

The PSID used a "cluster sample" when it started in 1968 in order to economize on interviewing costs. This design effect is typically a liability in statistical analyses because one has to account for non-independence across individuals within the same cluster. But for our purposes the clustering provides the unique opportunity to examine health outcomes for adults who were neighbors in 1968. Moreover, because all 1968 members within a given family are followed throughout their lives, we can examine the similarity in adult health status between spouses and unrelated individuals who were neighbors earlier in life.

Although the initial sample was clustered in 1968, because families and individuals have moved throughout the nation during the past 37 years, the total sample is now located in 2,570 different census tracts. The major advantage of the PSID for our purposes is that it offers the potential for constructing earlier-life neighborhood effects, coupled with rich detail on early-life and mid-life socioeconomic data, and data on dimensions of health status over time.

The selected sample consists of PSID sample members who were young adults when the study began and who have been followed into old age. Specifically, we choose PSID original sample members born between 1928 and 1948, which consists of adults who were in their 20s and 30s in the first wave of interviewing in 1968. We then obtain all available information on these individuals for each wave, 1968 to 2005. Therefore, by 2005 the oldest person in the sample is 77 and the youngest is 57.[1]

In the analyses, we define the neighborhood of residence during young adulthood as the census block where the respondent lived in 1968.[2] Thus, we are able to use a narrower, compact definition of neighborhood than the vast majority of previous studies of neighborhood effects. Typical studies use census tracts to define neighborhoods, and census tracts, which consist of roughly 5,000 families, are much larger than a census block which corresponds closely with a city block. In the original wave of the PSID, a represented census block contains 4 sampled families, on average.[3] The PSID cluster design is discussed in greater detail in Solon et al (2000).

Measurement of Health. The main adulthood health outcome examined is general health status (GHS), which was collected in the PSID in every wave since 1984. The general health status question is: "Would you say your health in general is excellent, very good, good, fair, or poor?" This question was asked of household heads and wives (if present) in each survey between 1984-2005, and was asked of all family members in 1986. GHS is highly predictive of morbidity measured in clinical surveys, and it is one of the most powerful predictors of mortality, even when controlling for physician-assessed health status and health-related behaviors. (For reviews of this extensive literature, see Idler and Benyamini (1997) and Benyamini and Idler (1999).) GHS is also frequently used as a global measure of health status and allows us to compare findings with those from related studies such as Case, Fertig, and Paxson (2005) and Currie and Stabile (2003).[4]

In order to scale the GHS categories, we use the health utility-based scale that was developed in the construction of the Health and Activity Limitation index (HALex). (A discussion of the various options for treatment of the GHS variable is described in Appendix B.) The HALex scores associated with GHS categories are based on the U.S. National Health Interview Survey, which contains a fuller health instrument than utilized in the PSID. A multiplicative, multi-attribute health utility model was used to assign scores and quantify the distance between the different GHS categories. The technical details of the scaling procedures are discussed at length elsewhere (Erickson, Wilson, Shannon, 1995; Erickson, 1998). Thus, using a 100-point scale where 100 equals perfect health, the interval health values associated with GHS used in this paper are: [95, 100] for excellent, [85, 95) for very good, [70,85) for good, [30,70) for fair, and [1,30) for poor health. Consistent with previous research, the skewness and nonlinearity of this scaling is reflected in the fact that the "distances" between excellent health, very good health, and good health are smaller than between fair and poor health. This scaling is currently used by the National Center for Health Statistics to estimate health-related quality of life measures and years of healthy life (*Healthy People* 2000). We then estimate all of the regression models of health status using the interval regression method. While the HALex approach with interval regressions is superior to alternatives, as described in the appendix, we have also estimated identical models to those reported in the tables but using poor/fair health as the dependent variable in a multi-level logit model.[5] The substantive conclusions are unchanged.

The sample includes males and females and all analyses control for gender, given well-known differences in health status, health behaviors, and labor market outcomes for men and women. Due to the complexity of the health status changes for women during the childbearing years, we exclude self-assessed health status measures of women in the years they were pregnant.

To increase the sample size as well as the proportion of poor and black families in the sample, we include both the Survey Research Center (SRC) component and the Survey of Economic Opportunity (SEO) component, commonly known as the "poverty sample," of the PSID sample. (See Hill (1992) for a detailed description of the two samples that compose the PSID.) We appropriately apply multi-level sample weights at the neighborhood and family levels to produce nationally-representative estimates.[6] The results are robust to the exclusion of the SEO sample, as estimates that exclude the SEO sample are nearly identical to those reported in the paper (results available upon request).

The resulting sample used to analyze general health status in adulthood contains 36,121 person-year observations from 2,730 individuals from 1,894 families, 1,457 neighborhoods, and 306 counties. The mean age is 53, with age ranging from 37 to 77, and an average of 13 observations per person. A total of 865 families contained married couples, and a total of 307 neighborhoods contained at least two different unrelated families.

The ability to conduct analyses within families and between neighboring families is a unique feature of the study. Because the study is among the first to report evidence of spousal correlations in health status, we include all neighborhoods to increase the effective sample size for the spousal correlation estimates. Results on the sub-sample that is restricted to neighborhoods containing adults from at least two different families yielded very similar magnitudes of spousal and young adult neighbor correlations in mid-to-late life health outcomes to those reported in the paper (results available upon request).

A key aspect of the data is that each individual is geocoded to the census block of residence and we utilize detailed information on neighborhood characteristics from respondent self-reports and neighborhood-level variables from the 1970-2000 Decennial Census. The self-reports of housing/neighborhood conditions include: whether live in Public Subsidized Housing; poor neighborhood for children, whether there exist plumbing problems, housing structural problems, security problems, cockroach or rat problems, insulation problems, neighborhood cleanliness problems, overcrowding, noise, or traffic problems, burglary, robbery, assault, drug use, or problems related to having too few police.[7]

Most prior studies that do use survey responses of neighborhood conditions characterize neighborhood features by relying on only a single individual's report of what happens in his or her neighborhood (and most often, even this information is not available). A more reliable approach to measuring neighborhood socioeconomic conditions is to aggregate the reports of multiple respondents living in the same neighborhood, which is the approach we use..[8] The self-reported survey information is used along with 1970-2000 census tract based measures—particularly, neighborhood poverty rate.

Our sample consists of original sample PSID individuals who were in young adulthood in 1968 (defined here as individuals in their 20s and 30s). The key measures of childhood socioeconomic status come from retrospective self-reports of childhood conditions collected in the early years of the PSID (which reduces measurement error because respondent reports are not as temporally distant from their childhood years). These measures include childhood poverty, parental education and occupation, year of birth, region of birth, and the type of community the individual grew up in. Ideally, we would also want to incorporate information on detailed measures of the neighborhoods in which individuals grew up, but this information is not available. We, however, make use of a unique set of measures in young adulthood, including crime, residential segregation, connectedness to informal sources of help, aspirations/motivation and long-term planning that were collected in the early years of the PSID. Appendix Table A0 lists the sources and years of all data elements along with details of the PSID survey questions used to construct these measures. Appendix Table A1 contains descriptive statistics for all childhood and young adult family and neighborhood measures for our sample by race.

VI. Econometric Model & Estimation Methods

In this section, we present an econometric model that illustrates the connections among spousal correlations, neighbor correlations, and regression analyses of neighborhood effects.[9] We begin by assuming the true model for health status is:

$$H_{sfn} = \alpha' X_{fn} + \beta' Z_n + \varepsilon_{sfn} \qquad (1)$$

where H_{sfn} denotes health status for spouse *s* in family *f* in neighborhood *n*, X_{fn} is the vector that includes all family characteristics (measured and unmeasured) that affect H_{sfn}, Z_n is the vector of all neighborhood characteristics that affect H_{sfn}, and ε_{sfn} is the error term that includes all individual-specific factors that are not related to X_{fn} or Z_n. Note that for illustrative simplicity, at this juncture, we do not attempt to incorporate dynamics and potential interactions between family and neighborhood background effects or nonlinearities into the model, but rather assume a linear representation.

Due to the self-selection of advantaged families sorting into advantaged neighborhoods for the reasons discussed in section III, we expect the family background factors, X_{fn}, and the neighborhood background factors, Z_n, to be positively correlated. Because it is difficult to fully and accurately measure every factor in X_{fn} and Z_n, the assumption that ε_{sfn} is uncorrelated with the observable measures of X_{fn} and Z_n will be violated, leading to biased estimates of neighborhood effects (β) and family background effects (α). Using the taxonomy of Manski (1993), it is not possible to distinguish the two types of "social effects" -- "endogenous effects" and "exogenous effects" -- from the nonsocial "correlated effects". Manski also demonstrates that it is not possible to distinguish the two types of social effects from each other.

Therefore, the first goal of the analysis is focused on an overall assessment of the relative contributions of individual, family and neighborhood influences during childhood and young adulthood on health in mid-to-late life. We then analyze the relative contribution of a rich array of measured childhood socioeconomic conditions, young adult family and neighborhood, and individual covariates to the total variation from each component, and test hypotheses about the effects of specific characteristics of families and neighborhoods.

The strategy for assessing the importance of contextual effects involves estimating the fraction of variation in health outcomes of interest that lies between families and neighborhoods, to provide an upper bound on the possible effect of these contexts. The intuition motivating the use of this strategy is that if family and residential community are important determinants of health outcomes, there will be a strong correlation between spouses in their health outcomes, as compared to two arbitrarily chosen individuals. And if the neighborhood where the individual lived as a young adult has an enduring effect or predicts future health, it will show up as a strong correlation between neighboring adult's subsequent later-life health outcomes.

As demonstrated in Solon et al (2000), using the additive model of the effect of family and neighborhood context in equation (1), the population variance of H_{sfn} can be decomposed as:

$$Var(H_{sfn}) = Var(\alpha' X_{fn}) + Var(\beta' Z_n) + 2Cov(\alpha' X_{fn,} \beta' Z_n) + Var(\varepsilon_{sfn}) . \quad (2)$$

Similarly, the covariance in H_{sfn} between spouse s and s' is:

$$Cov(H_{sfn}, H_{s'fn}) = Var(\alpha' X_{fn}) + Var(\beta' Z_n) + 2Cov(\alpha' X_{fn,} \beta' Z_n) . \quad (3)$$

The spousal correlation, $\text{cov}(H_{sfn}, H_{s'fn}) / \text{var}(H_{sfn})$, measures the proportion of the total variation in the health outcome under consideration due to factors shared by married couples. From (3) we see that spouses have correlated health outcomes because they have shared family and neighborhood backgrounds, corresponding to the first and second terms of (3), respectively. The sorting of families into neighborhoods is reflected in the third term. The spousal covariance then captures all measured and unmeasured factors shared by married couples that may have an impact on health outcomes, such as adult socioeconomic status, family structure, similar childhood socioeconomic status due to assortative mating patterns along socioeconomic lines, as well as neighborhood effects stemming from the quality of neighborhood conditions.

Augmenting the estimation of spousal correlations with the estimation of young adult neighbor correlations in later-life health enables us to bound the relative importance of young adult family and neighborhood factors. To see this, note the covariance between neighbors is:

$$Cov(H_{sfn}, H_{s'f'n}) = Cov(\alpha' X_{fn}, \alpha' X_{f'n}) + Var(\beta' Z_n) + 2Cov(\alpha' X_{fn,} \beta' Z_n) \qquad (4)$$

The last two terms in (3) and (4) are identical, so we expect the covariance between neighbors to be smaller than the covariance between spouses because married couples share both the same adult neighborhood and family. If the covariance among neighbors is small relative to the covariance among spouses, the family effects, which are represented by the first term in (3), must be the main source of the covariance among married couples.

The neighbor correlation, $\text{cov}(H_{sfn}, H_{s'f'n}) / \text{var}(H_{sfn})$, measures the proportion of the variation in the health outcome that can be attributed to factors shared by individuals from the same neighborhood. In (4), we notice that the neighbor covariance consists of more than the variance in (effect-weighted) neighborhood characteristics given in the second term, and it should therefore be viewed as an upper bound of the neighborhood influence on the covariance in H_{sfn} between neighbors. The first and third terms are both expected to be positive, leading to an upward bias. The first term represents the sorting of similar families into the same neighborhoods, since neighbors share similar family characteristics. Similarly, the third term also represents sorting, in that it captures sorting of disadvantaged families into disadvantaged neighborhoods. We see that positive sorting, $Cov(\alpha' X_{fn}, \alpha' X_{f'n}) \geq 0$ and $Cov(\alpha' X_{fn,} \beta' Z_n) \geq 0$, implies that $Var(\beta' Z_n) \leq Cov(H_{sfn}, H_{s'f'n})$.

Access to neighborhood identifiers and family characteristics in the same data enables us to tighten the upper bound on the neighborhood effects and also establish a lower bound on the family effects. First, it follows from (4) that the upper bound on the neighborhood effects can be made tighter by introducing observable family characteristics shared by the neighbors, and by subtracting that as an observable part of the first term of (4). Following Solon et al (2000) and Altonji (1988), we estimate the part of $\alpha' X_{fn}$ related to observable young adult family characteristics such as income, education, family structure, race, health insurance coverage, alcohol and cigarette use, housing quality, and aspirations. Let $\widetilde{X}_{fn}$ denote the observable subset of family characteristics with associated parameters $\hat{\alpha}$ estimated *within* neighborhoods. We then subtract off the sorting component arising from the fact that similar families tend to cluster in neighborhoods,

$$Cov_{adj}(H_{sfn}, H_{s'f'n}) = Cov(H_{sfn}, H_{s'f'n}) - Cov(\hat{\alpha}'\widetilde{X}_{fn}, \hat{\alpha}'\widetilde{X}_{f'n}) \; . \qquad (5)$$

The tighter upper bound on neighborhood effects also implies a tighter lower bound on family effects. Specifically, the difference between the spousal correlation and the adjusted neighbor correlation represents a lower bound of the magnitude of the family effect on the health outcome of interest. We refer to this as the "adjusted spousal correlation."

Four-Level Hierarchical Random Effects Interval Regression Model. We decompose both the variance of the level of health and the rate of health depreciation over time into the fraction that lies between neighborhoods, families, and individuals. In order to decompose both the total variation in the health level and the health depreciation rate, we estimate a four-level hierarchical random effects interval regression model. The data are hierarchical because we have multiple observations over time of individuals who are nested within families, which are nested within neighborhoods and counties. Multilevel modeling techniques can accommodate the hierarchical and unbalanced structure of our data, non-independence of the (sometimes overlapping) pairs of spouses and neighbors, as well as the non-normality of health (Raudenbush and Bryk, 2002).

We begin by estimating the four-level hierarchical random effects model[10] given by

$$H^*_{tsfn} = (\beta_{0000} + \beta_{1000} * Age_t) + (\eta_{000n}) + (\phi_{00fn}) + (\delta_{0sfn}) + \varepsilon_{tsfn} \qquad (6)$$

We estimate these models separately at middle age (ages 35-55) and later-life (ages over 55), in order to gain greater insight into the extent to which family and neighborhood factors earlier in life influence the trajectory of health in mid-to-late life. These unconditional baseline models also include controls for year of birth and quadratic terms for age (suppressed in the above notation for illustrative simplicity). The indices *t*, *s*, *f*, and *n* denote time, individuals, families, and neighborhoods, respectively, where there are

$t = 1, 2, \ldots, O_{sfn}$ observations over time of individual *s* in family *f* in neighborhood *n*;

$s = 1, 2, \ldots, S_{fn}$ spouse in family *f* in neighborhood *n*;

$f = 1, 2, \ldots, F_n$ families in neighborhood *n*;

$n = 1, 2, \ldots, N$ neighborhoods.

The neighborhood-, family-, and individual-level random effects capture unobserved characteristics of the neighborhood, family, and individual. The neighborhood random intercept coefficient is represented by η_{000n}; the family random intercept coefficient is represented by ϕ_{00fn}; the individual random intercept coefficient is represented by δ_{0sfn}; and ε_{tsfn} represents the individual transitory component of self-reported health (which includes measurement error). Each of these random effects are assumed to be normally distributed with a mean of 0, and var(η_{000n})=σ^2_{0n}, var(ϕ_{00fn})=σ^2_{0fn} , var(δ_{0sfn})=σ^2_{0sfn} , and var(ε_{tsfn}) $=\sigma^2_{tsfn}$. Age_t is the individual's actual age at time *t* centered around the mean age in the sample. Since neighborhoods are nested within counties, we also estimated five-level hierarchical models, where the hierarchical levels represented counties, neighborhoods, families, and individuals over time. Those models were estimated as a robustness check to ensure that the young adult neighborhood random effects components were not primarily driven by effects operating at higher geographic levels of aggregation. However, those models did not significantly improve the fit and the between-county random effects component was not statistically significant, which supports the use of the four-level hierarchical model reported throughout the paper. All standard errors are Huber-corrected, clustered on county.

Health varies with age and gender. Because we did not want our estimates of spousal and neighbor correlations to reflect the influence of either of these two demographic factors, we adjusted for them in our baseline model by including gender and a quadratic specification of age as explanatory variables.

Of primary interest is the decomposition of the variance of the level of health in mid-to-late life into their within-family, between-family within-neighborhood, and between-neighborhood components. In this model, individuals from the same neighborhood but not in the same family (i.e., neighbors) are correlated because they share the random effect η_{000n}, and married couples are correlated because they share the random effects η_{000n} and ϕ_{00fn}. Here we want to evaluate the health correlation between spouses at the same age, and evaluate the health correlation between neighbors at the same age. In this model, the spousal correlation and neighbor correlation in the level of health can be computed, respectively, as:

$$\rho_{\text{spousal,healthlevel}}(\text{age}) = \frac{(\sigma_{0n}^2) + (\sigma_{0fn}^2)}{(\sigma_{0n}^2) + (\sigma_{0fn}^2) + (\sigma_{0sfn}^2)}$$

$$\rho_{\text{neighbor,healhlevel}}(\text{age}) = \frac{(\sigma_{0n}^2)}{(\sigma_{0n}^2) + (\sigma_{0fn}^2) + (\sigma_{0sfn}^2)}$$

The spousal correlation is between H_{sfn}^* and $H_{s'fn}^*$, evaluated at the same age; the neighbor correlation is between H_{sfn}^* and $H_{s'f'n}^*$, evaluated at the same age. Our interest is in the permanent (rather than the transitory) component of health, so we do not include the temporal variation of health in the denominator.

We can then use the estimated spousal and young adult neighbor correlations in health at mid-life and later-life to construct an age-profile of spousal and neighbor health correlations. This will enable us to better assess the later-life health consequences of neighborhood and family influences experienced earlier in the life cycle.

Estimating "Adjusted Neighbor Correlations". We estimate "adjusted neighbor correlations", which are net of the similarity arising from young adult neighbors having similar observed family characteristics. To extract the impact of similar family characteristics out of the neighbor correlation, we first estimate the following regression, where for ease of exposition we omit the random effects terms that are included in the estimated model:

$$H^{*}_{tsfn} = \alpha_0 age_{tsfn} + \alpha_1 gender_{sfn} + \alpha_2{}' X_{\bullet\bullet fn} + \alpha_3{}' (\overline{X_{\bullet\bullet\bullet n}}) + \varepsilon_{tsfn} \,, \qquad (7)$$

where $X_{\bullet\bullet fn}$ is a vector of young adult family characteristics including: average annual family income-to-needs ratio (based on the five-year average as reported in 1967-1972), education, family structure, race, health insurance coverage (as reported in 1967-1972), annual expenditures on cigarette and alcohol consumption (based on the five-year average in 1967-1972), housing plumbing and insulation problems, connectedness to informal sources of help, and aspirations. $\overline{X_{\bullet\bullet\bullet n}}$ is a vector of the 1968 neighborhood-level means of the same above variables.

Inclusion of family-level and neighborhood-level variables measuring the same concepts enables the vector α_2 of coefficient estimates to capture the within-neighborhood effects of these family characteristics. Using the within-neighborhood estimates of the family effects of income, education, race, family structure, health insurance coverage, health behaviors, connectedness to informal sources of help and housing quality on health in adulthood, will ensure the coefficients

(α_2) will not be biased by omitted neighborhood variables. This follows from the fact that the neighborhood-level unmeasured factors can only be correlated with the neighborhood-level mean of the covariates. In combination, the resulting estimates of the effects of these family characteristics can be taken as a conservative estimate of $\alpha' X_{fn}$ in equation (1).

We then estimate the between-neighborhood variance in $\hat{\alpha}' X_{fn}$ by estimating a hierarchical random effects model of $\hat{\alpha}' X_{fn}$ on neighborhood-level, family-level, and individual-level random effects. We then subtract the estimate of the between-neighborhood variance in $\hat{\alpha}' X_{fn}$ from the estimate of the overall between-neighborhood variance in H^*_{sfn}. Dividing the resulting quantity by $V\hat{a}r(H^*_{sfn})$ yields a tighter upper bound on the proportion of $Var(H^*_{sfn})$ that can be attributed to neighborhood effects.

The estimates of "adjusted neighbor correlation" enable us to ascertain how much of the raw neighbor correlation is due to young adult neighbors having similar (observable) family characteristics. We then investigate to what extent observable childhood socioeconomic status and young adult neighborhood- and family-level characteristics explain the observed spousal and neighbor correlations at middle age and late-life.

We provide estimates of the distinct effects of neighborhood and family level background variables measuring the same concepts—for example, the effects of family SES conditional on neighborhood SES and vise versa. In addition, explicitly measuring the magnitude of variation in the effects of unmeasured factors allows an assessment of the importance (quasi-R^2) of the measured variables, *X,* in total variation at each level (e.g., measures vs. unmeasured neighborhood characteristics). In a final set of models, we include contemporaneous measures of socioeconomic status in adulthood into the four-level hierarchical random effects model to examine the extent to which the resemblance of young adult neighbors' subsequent health in adulthood may be due to the similarity of their economic status in adulthood. These estimates are only suggestive because of the well known endogeneity between contemporaneous health and SES. The results serve to demonstrate what aspects and sources of later-life health disparities may be missed using traditional models that focus more on contemporaneous socioeconomic factors, without the emphasis on earlier life factors.

Family income and neighborhood poverty are dimensions of family and neighborhood background that we give particular emphasis to in the regression analysis. Living in a neighborhood with concentrated poverty may have grave consequences above and beyond those of growing up in a poor family because of social isolation, crime, weakened social institutions, unrelenting stress, inferior health care accessibility, and other factors. We make use of a unique set of measures in young adulthood, including crime, residential segregation, connectedness to informal sources of help, aspirations/motivation and long-term planning that were collected in the early years of the PSID. These factors may themselves be the product of living in a high poverty neighborhood and may represent important pathways through which exposure to depressed neighborhood environments earlier in life affect health trajectories later in life. However, controlling for this myriad of ways in which those who reside in high poverty neighborhoods may differ from individuals who live in affluent neighborhood environments allows one to generate a more conservative estimate of the effect of neighborhood poverty itself, as well as shed light on the factors that affect adult health status.

Sensitivity Analysis. We conduct a sensitivity analysis to test the robustness of the estimated effects of young adult neighborhood poverty to selection bias due to an omitted variable. The goal is to assess how the point estimate and confidence interval of the effect of neighborhood poverty change under the presence of selection bias of varying strengths. We use a novel empirical approach, recently proposed by Altonji et al. (2005) and Krauth (2006), to perform the sensitivity analysis. This analysis allows one to determine the threshold of selection on unobservables, if any, at which neighborhood poverty during young adulthood no longer has a significant effect on adult health. The approach uses the statistical relationship between observed explanatory variables in a regression as a guide to generate plausible estimates about the relationship between observed and unobserved variables. The sensitivity parameter, θ, can be defined as

$$corr(X_k, u) = \theta corr(X_k, X\beta - X_k\beta_k) ,$$

where θ indexes the magnitude of the correlation between observables and unobservables relative to the analogous correlation among observables themselves. In other words, the correlation between the neighborhood poverty rate and the (effect-weighted) unobservables is proportional to the correlation between the neighborhood poverty rate and the effect-weighted observables. The standard exogeneity assumption is the special case of $\theta=0$. This approach provides a way to construct bounds on the effect of neighborhood poverty during young adulthood on mid-to-late life health based on the bounds one is willing to place on the sensitivity parameter θ (i.e., the relative correlation).

Altonji et al. (2005) argue that if the observable determinants of an outcome are truly just a random subset of the complete determinants, selection on observable characteristics must be equal to selection on unobservable characteristics. Because the PSID was conducted specifically to study family factors that affect well-being, we would expect selection on observable factors to be greater than selection on unobservable factors; in other words, the extensive measures of family and neighborhoods captured in the PSID are likely to be the most important determinants of adult health. Thus, estimates obtained under the assumption of equal selection will be biased downwards.

VII. RESULTS

Before discussing the estimates from the hierarchical models, we begin by presenting nationally-representative estimates of the bivariate relationship between mid-to-late life health status and socioeconomic status in childhood (i.e., parental education and poverty status), socioeconomic status in young adulthood (i.e., poverty status and own educational attainment), and neighborhood quality in young adulthood (i.e., poverty and crime, race and residential segregation, and neighborhood housing quality). The unadjusted spousal and young adult neighbor correlations of health in mid-to-late life are presented next. We then examine how much of the adult neighbor correlations can be explained by the fact that families in a neighborhood tend to be similar as opposed to emanating from neighborhood effects *per se*. Then, we estimate the magnitude of the effects and attempt to explain the life-cycle pattern of spousal and neighbor correlations and explore potential mechanisms that underlie the relative roles of neighborhood and family factors on the health trajectory over the course of adulthood.

Descriptive Results. Figures 1-5 display the proportion of years in poor health as an adult as well as the age pattern of the health index (which was described earlier). The age patterns of the conditional expectations are calculated using a Jianqing Fan (1992) locally weighted regression smoother, which allows the data to determine the shape of the function, rather than imposing, for example, a linear or quadratic form. The differences presented are all statistically significant.

Nearly one-quarter of adulthood years between ages 40 and 65 is spent in fair or poor health among those who grew up in poverty, while that rate is cut in half (13%) among those who grew up in non-poor families (Figure 1). Similarly, among individuals whose parents lacked a high school education, roughly one-quarter of adulthood years between ages 40 and 65 is spent in fair or poor health, whereas these rates among those whose parents attended high school but not college, and those who had college-educated parents, were about twelve percent and eight percent, respectively.

Nearly forty percent of adulthood years between ages 40 and 65 is spent in fair or poor health among high school dropouts; in comparison, these rates for high school graduates, those who attended college (without earning a BA/BS degree), and college graduates were eighteen percent, fourteen percent, and six percent, respectively. The bivariate relationship between the family income-to-needs ratio as a young adult and mid-to-late life health exhibit nonlinearities, with those in poverty as young adults experiencing significantly higher rates of problematic health throughout adulthood (Figure 2). The health deterioration rate does not appear to be significantly more rapid in adulthood among those who had low family socioeconomic status as young adults (Figure 2). In contrast, as shown in Figures 3 and 4, racial health disparities and the socioeconomic gradient in health along neighborhood quality dimensions are not only large, but they also appear to widen over the course of adulthood, as the health deterioration rate is more rapid in adulthood among those who lived in more disadvantaged neighborhood environments as young adults.

Individuals who lived in high crime neighborhoods as young adults have worse health during mid-to-late life, relative to those who lived in low crime neighborhoods. In particular, nearly one-quarter of adulthood years between ages 40 and 65 were spent in fair or poor health, on average, among individuals who lived in high crime neighborhoods as young adults, compared with seventeen percent among those who lived in neighborhoods without crime problems. Similar patterns of differences exist in adulthood health between individuals who lived in environments with neighborhood plumbing and insulation problems, relative to individuals who resided in environments that did not have these problems as young adults (Figure 5).

Unadjusted Spousal and Young Adult Neighbor Correlations in Later-life Health. Table 1 presents the estimates from the baseline four-level hierarchical random effects models, separately at ages 35-55 and at ages over 55—these models control only for age, gender, and year of birth. As shown in Table 1, the random effects estimates are all significant at each of the young adulthood neighborhood, family and individual levels.

The baseline random effects models in Table 1 enable the measurement of the overall magnitude of variation in mid-to-late life health related to young adult neighborhood, family, and individual-level factors. The spousal and neighbor correlation estimates are based on the decomposition of variance over time into the fraction that lies between neighborhoods, families, and individuals. The age-profile of the estimated unadjusted spousal and neighbor correlations calculated from the baseline models are summarized in the first row of Table 5. We find that the spousal correlation in general health status is 0.46, on average, across ages 40-70, and declines somewhat during ages after 55 (though some of this decline may be related to selective mortality—see the Appendix for more discussion).

Johnson (2008) used the PSID and analyzed health outcomes in childhood and adulthood through mid-life of individuals born between 1951-1970 followed up to 2005. His results show sibling correlations are large throughout at least the first 50 years of life: the sibling correlation in general health status is roughly 0.6—suggesting that three-fifths of adult health disparities may be attributable to family and neighborhood background. The comparison of sibling correlation in adult health with spousal correlation in adult health serves as an instructive way to explore the relative impact of shared household environment versus shared genetic factors using relatedness of individuals in the household (after accounting for cohort and age differences across the samples).[11] The spousal correlation estimates that correspond with the similar ages as the prior evidence in Johnson (2008) on sibling correlation of health in mid-life compare favorably: 0.47 for the spousal correlation and 0.63 for the sibling correlation.

Spousal correlations by themselves cannot disentangle how much of the resemblance among their health outcomes is due to the effects of shared household environment and family background and how much is due to the effects of neighborhoods. Augmenting the spousal correlation estimates with corresponding neighbor correlation estimates reveals young adult neighborhood and family factors are both important determinants of general health status in mid-to-late life. While the young adult neighbor correlations are smaller than the spousal correlations, they are significant through middle-age and old age. In particular, the young adult neighbor correlation in health at ages 40-70 averages 0.33. These magnitudes are particularly noteworthy given the fact that the vast majority of these individuals no longer live in the same neighborhoods they once shared as young adults (in this paper young adulthood is defined as individuals in their 20s and 30s). We will return to the issue of the extent of residential mobility and persistence in neighborhood quality across the life course, and its implications on these results in the final section.

Adjusted Spousal and Young Adult Neighbor Correlations in Later-life Health. From the adjusted neighbor correlation estimates, we find that observable family sorting (controlling for a broad array of young adult family characteristics including income, education, race, family structure, health insurance coverage, health behaviors, connectedness to informal sources of help and housing quality) explain some but not all of the resemblance in adulthood health status among individuals who lived in the same neighborhood as young adults. (The full model results used to compute the adjusted neighbor correlations are not shown to conserve space, but are available from the authors upon request). Specifically, the adjusted neighbor correlation is roughly 20% lower than the unadjusted neighbor correlation (Table 5, row 2). These results imply that differences in neighborhood quality during young adulthood may account for up to roughly one-quarter of health disparities in mid-to-late life.

These results build directly on the previous findings of Johnson (2008), which illuminated the importance of family background and neighborhood origins on child health and adult health through mid-life. This prior evidence facilitates the identification of the antecedents of health at mid-life and provides us with a better understanding of the early risk factors for health decline among older adults. In particular, Johnson's (2008) previous estimates of childhood neighbor correlations in early-to-mid life health suggest that disparities in neighborhood background account for 35-40 percent of the variation in health status in mid life. These composite effects emanate from the direct effects of neighborhood quality during childhood on child health that may carry over into adulthood, as well as indirect effects via the economic mobility process. Moreover, differences in developmental health trajectories explain much of the variance in the nature and rate of later declines in health. One aim of the comparison of the childhood neighbor correlations in adult health with the adult neighbor correlations in health among the elderly estimated in this paper is to contribute to our understanding of how neighborhood effects vary over the life cycle.

While the implied scope of young adult neighborhood influences on subsequent health status is sizable, the estimated adjusted young adult neighbor correlations are smaller than previously estimated child neighbor correlations in mid-life health. In particular, net of the similarity arising from similar family characteristics, child neighbor correlations in mid-life health average about 0.40, as compared with the (adjusted) young adult neighbor correlations in mid-life health of 0.27. This evidence suggests a potentially more prominent role of childhood neighborhood factors than neighborhood environments in young adulthood. It is important to bear in mind, however, that the life-course conditions and exposures of individuals born into socioeconomically disadvantaged families and neighborhoods may differ by birth cohort; as a result, even though our analyses control for year of birth, these estimates are not directly comparable (given that the sample used to estimate child neighbor correlations consists of individuals born between 1951 and 1970, whereas our older birth cohort sample consists of individuals born between 1928 and 1948).

It is important to note that, without access to nationally-representative longitudinal data and the ability to identify the permanent component of health, our transitory component would have been captured in resultant point-in-time estimates, significantly diluting the relevant estimated spousal and neighbor correlations in adult health that are of primary interest. This result demonstrates the importance of correcting for measurement error, transitory fluctuations and unrepresentative homogenous samples, and parallels those found in the earlier literature that focused on the permanent component of adult earnings (Solon *et al*., 1991).

Magnitude of effects and potential pathways. What do these correlation estimates mean in terms of the absolute size of the effects of family and neighborhood factors, and what aspects of family socioeconomic status and young adult neighborhood matter to explain the lasting importance of earlier life conditions on late-life health? Estimates of the neighborhood random components (σ_n) indicate that neighborhood quality during young adulthood has significant and enduring effects on general health status over the course of adulthood. It is also possible that some of this underlying relationship may be due to differences in childhood neighborhood quality and the correlation between childhood neighborhood and young adult neighborhood factors. (We cannot directly investigate this issue for the older birth cohorts analyzed in this paper because we do not have information on detailed neighborhood-level factors experienced during childhood).

With these questions in mind, we next estimate a series of models building toward a full model specification that includes a rich array of observable child socioeconomic status measures and young adult family-level and neighborhood-level characteristics to attempt to identify earlier life determinants of health in mid-to-late life. We conduct a systematic analysis of socioeconomic and racial health disparities in adulthood, and attempt to explain the level and age-profile of spousal and neighbor correlations. Tables 2-4 contain the regression results for all adulthood years (ages 35+), and separately for ages 35-55, and ages older than 55—where the series of models reported include the raw age-adjusted race gap (column(2)), a model that includes controls for childhood socioeconomic characteristics (column(3)), a model that controls for both childhood SES and young adult family characteristics (column (4)), a model that controls for childhood SES, and both young adult neighborhood and family characteristics (column (5)), and a final model that controls for childhood SES, young adult neighborhood and family characteristics, and contemporaneous socioeconomic factors (column (6)). To conserve space, we will integrate the discussion of the results contained in Tables 2-4 and summarize together the sequential set of hierarchical random effects models estimated.

Separately identifying the causal pathways through which earlier life socioeconomic factors influence late-life health is an arduous task, and one must caution against drawing causal inferences from these coefficient estimates. The estimates are intended instead to summarize the relationships between the health trajectory in adulthood with various dimensions of neighborhood and family background.[12] Moreover, distinguishing between latent, cumulative, and pathway process that may involve dimensions of both families and neighborhoods, as discussed in Section II, is beyond the scope of the present paper, but remains an important area for future research.

Gaps in health between blacks and whites are large and exist at all stages in life. As shown in column (2) of Tables 2-4, respectively, the general health status (GHS) index in adulthood, on average, is 13.9 points lower for blacks and this gap increases in levels and in proportionate terms in adulthood. A useful way to interpret the estimate is in relationship to the size of the effect of age on health, with the race gap in ages over 55 equivalent to blacks (on average) reaching a level of health deterioration about 30 years prior to their white counterparts. That is, GHS is 13.9 points lower for black adults (column (2) of Table 2), which is equal to roughly 30 years evaluated at an effect of age of -0.45.

As evidenced in column (3) of Tables 2-4, a few childhood SES measures account for 30-35 percent of the black-white gap during adulthood.[13] Furthermore, part of the resemblance in later-life health between spouses and young adult neighbors are linked to the similarity of their childhood SES conditions. Conditional on childhood SES factors, the spousal correlation is 0.36 and the young adult neighbor correlation in later-life health is reduced to 0.2, which is roughly twenty-five percent lower than the adjusted neighbor correlations (see the third row of Table 5). Most prominent among these factors is childhood poverty and parental education, which in turn influence young adult socioeconomic status attainments and thereby may affect access to neighborhood amenities in adulthood. For example, the results in column (3) of Table 4 indicate that growing up poor is associated with a 3.75 lower health status index at ages over 55, which is equivalent to 8.3 years older. In addition, both mother's and father's education are strongly associated with adult health status; the estimated effect of father's education becomes smaller than the corresponding effect of mother's education after controlling for poverty status. In particular, having a mother who dropped out of high school is comparable to the estimated childhood poverty effect and is associated with 4.3 lower adult health status index (relative to having a mother who was a high school graduate), while having college-educated parents is associated with beneficial health impacts in adulthood. Parental occupation is more weakly associated with adult health after accounting for parental education. The type of community the individual grew up in and the region of birth are each also shown to be strongly associated with adult health—in particular, individuals who grew up in large, urban areas (relative to small towns or on a farm) experienced better health in adulthood, and those born in the South exhibited significantly worse adult health than those from other regions of the country.

Our data contain a richer set of young adult family and neighborhood attributes. Once we include both the set of childhood SES and young adult family measures, the black-white difference in health status at ages over 55 is reduced to one-third of its original raw age-adjusted gap (see column(4) in Table 4). Most prominent among these young adult family factors is family income, with substantially larger impacts of income in the lower tail of the distribution highlighting the negative effects of poverty. For example, the results in column (3) of Table 6 indicate that a one-unit increase in the family income-to-needs ratio (during young adulthood) from half of the poverty line to 1.5 times the poverty line translates into a 8.5 point increase in adult GHS at ages over 55 (0.5*14.4338+0.5*2.6004), which is equivalent to 18.9 years younger.

One's own educational attainment, health insurance coverage, cigarette smoking and alcohol use in young adulthood are each also shown to be strongly associated with later-life health.

The estimates in column (5) of Tables 2-4 show that once we control for child SES and both young adult neighborhood and family factors, the black-white difference in health status at ages over 55 is reduced by 76 percent. The coefficients on the childhood SES measures are reduced significantly with the inclusion of the set of young adult family and neighborhood factors, but the childhood factors remain large and significant (comparing columns (4) and (5) in Tables 2-4).

Most salient among the young adult neighborhood factors is concentrated neighborhood poverty. To put the magnitudes in perspective, it is useful to consider that the estimated later-life health differences between those who lived in a medium poverty neighborhood versus a low poverty neighborhood during young adulthood are on par with the estimated impacts of smoking as a young adult. As shown in column (5) of Table 4, the results for health at ages over 55 show that a ten percentage-point increase in the neighborhood poverty rate (in young adulthood) from 10 to 20 percent is related to a 1.49 reduction in GHS, and living in a high poverty neighborhood corresponds with a 6.34 lower GHS score in later life, relative to being living in a low poverty neighborhood as a young adult. This latter effect is equivalent to reaching a level of health deterioration roughly 14 years sooner for an individual who lived in a high poverty neighborhood, relative to the simulated health trajectory experienced by one who lived in a low poverty neighborhood earlier in adulthood. For purposes of causal inference, the robustness of this result to alternative thresholds of selection on unobservables is analyzed in the following section. For comparison purposes, we note that while the estimated impacts of neighborhood poverty in young adulthood on subsequent health presented in this paper are substantive, they are noticeably smaller than the impacts of childhood neighborhood poverty on adult health reported in Johnson (2008). This may be due to the fact that childhood SES factors are more closely linked with school quality and opportunities that affect socioeconomic mobility prospects which in turn have far reaching impacts on later-life health.

Several other related dimensions of neighborhood disadvantage experienced in young adulthood had substantive, independent influences on the health trajectory over the course of adulthood, including high crime, residential segregation, neighborhood housing problems, neighborhood connectedness to informal sources of support (which may serve as a proxy for social cohesion), and neighborhood-level average aspirations for socioeconomic attainment. These factors appear to have stronger relationships with health over time, with stronger links to health at ages over 55 relative to middle-age. The age-profile of these estimated effects suggest that the linkages may be the result of cumulative exposure to disadvantaged environments taking a toll on health later in life that may be reinforced with how these factors earlier in life influence the socioeconomic mobility process. For example, as shown in column (5) of Table 4 for health status at ages over 55, living in a high crime neighborhood as a young adult reduces GHS by 1.3 points; a 25 percentage-point increase in childhood residential segregation (dissimilarity index increase from 50 to 75) is related to a 1.2 point reduction in GHS (similar effects for both blacks and whites (results not shown)). Taken together, the cumulative set of childhood SES measures, and young adult neighborhood and family factors account for fifty-four percent of the neighborhood-level variance at ages over 55 (implied quasi-R^2 at the neighborhood level). After controlling for observable childhood and young adult SES, the similarity of both spouses' and young adult neighbors' later-life health outcomes are less marked. Namely, conditional on childhood SES, and young adult neighborhood and family factors, the spousal correlation is 0.28

and the young adult neighbor correlation in mid-to-late life health is reduced to 0.13, which is roughly half the size of the adjusted neighbor correlations (see the third row of Table 5).

Most of the neighborhood effects literature that has examined health outcomes has investigated whether contemporaneous neighborhood factors are significantly associated with health at a point in time (see, for example, literature review articles by Ellen et al., (2001) and Morenoff and Lynch (2002). In column (6) of Tables 2-4, we use our models to re-examine this issue. The final model includes both the full set of childhood SES and young adult neighborhood and family factors along with the contemporaneous adult neighborhood poverty rate (differences in the functional form specified for contemporaneous neighborhood poverty yielded nearly identical results). As shown in column (6) of Tables 2-4, the results reveal that while three-quarters of racial differences in health at ages over 55 can be accounted for by childhood SES and young adult family and neighborhood factors, contemporaneous neighborhood factors independently account for little of this gap. Contemporaneous adult neighborhood poverty was only weakly related to health status in mid-to-late life. The coefficient estimates on the childhood SES and young adult neighborhood and family factors remain significant and are not significantly reduced with the inclusion of the contemporaneous adult neighborhood poverty rate.

Table 5 presents the spousal and young adult neighbor correlations in mid-to-late life health status as estimated from these hierarchical random effects models, for the unconditional, adjusted, and conditional model estimates after controlling for childhood factors and young adult adult socioeconomic status. The broad array of available measures of family and neighborhood background, which are in many ways unique to the PSID, is a tremendous asset to the analyses.

Sensitivity to Selection Bias. The estimates of the significant effects of neighborhood poverty during young adulthood on later-life health reported in Table 4 are based on models in which exogeneity is assumed. Concentrated neighborhood poverty is associated with significantly elevated risks of problematic health, even after controlling for an extensive set of individual and family-level factors. We next evaluate the robustness of these results to deviations from exogeneity. Table 6 presents the range of estimated coefficients and standard errors on young adult neighborhood poverty as a function of the ratio of selection on unobservables to selection on observables. We find that the effect of young adult neighborhood poverty on health status later in life remains large and significant even with a reasonably large amount of selection on unobservable factors. Even if the correlation between neighborhood poverty and unobserved outcome-relevant factors was assumed to be equal to the correlation between neighborhood poverty and observed-relevant factors, this does not eliminate the significant effect of neighborhood poverty during young adulthood on health status later in life. In other words, even if one assumes a significant degree of selection on unobservables, one cannot attribute the entire effect of neighborhood poverty to selection bias.

While there is no single perfect solution to address the endogeneity of residential location, there are additional tests we employed to determine if selection bias is driving the results. In additional model results not shown, we find that the strength of the estimated effects of neighborhood poverty during young adulthood vary with the number of known neighbors and duration of exposure to concentrated neighborhood poverty; this result is not consistent with a simple sorting on unobservables story to explain the findings. The fact that the effect of concentrated neighborhood poverty is weaker when the duration of exposure is brief and individuals know few of their neighbors suggests that selection bias is not driving these results.

If effects simply represented unmeasured family factors, then the number of years in the neighborhood and the number of neighbors known by name should not be associated with the strength of these effects. But that is not the case here. These results are not consistent with the hypothesis that families predisposed to poor outcomes select into poor neighborhoods, but the pattern of results, taken together, point towards true causal effects of concentrated neighborhood poverty, rather than results purely driven by selection effects.

VIII. CONCLUSION

Previous research demonstrates that health status in the US varies strongly across local, state, and regional settings (Murray, Michaud, McKenna, and Marks, 1998). This paper investigates multilevel causation of these differences over the course of adulthood and provides insights into how family and neighborhood environments earlier in life influence the health trajectories of individuals in ways that cannot be reduced to the characteristics of the individuals themselves.

In this paper, we have used correlations based on a nationally representative longitudinal sample of married couples and neighbors followed from young adulthood through elderly ages to estimate bounds on the possible causal effects of young adult family and neighborhood characteristics on general health status in mid-to-late life. Estimates based on four-level hierarchical random effects models consistently show a significant scope for both young adult family and neighborhood factors, though a noticeably smaller scope as compared with childhood family and neighborhood influences illuminated in prior work (Johnson, 2008). One must bear in mind, however, that the adult cohorts analyzed across the two studies are not directly comparable due to differences in birth cohort and the age distribution of the adults. The estimates suggest that disparities in neighborhood conditions experienced in young adulthood account for up to one-quarter of the variation in health status in mid-to-late life. While the neighbor correlations must be strictly interpreted as upper bounds, the estimates suggest that neighborhood factors experienced earlier in the life course influence both contemporaneous and future health outcomes.

Research on how neighborhood and family background influence later-life health is one with potential endogeneity issues that are not amenable to the usual microeconometric corrections through use of instrumental variables or fixed effect approaches, and for which the extant experimental evidence is likely too short a time horizon to detect effects on overall health status. Instead of attempting to remove or avoid selection bias caused by unobserved factors, the methods employed in this paper assess how the presence of varying levels of selection bias would alter conclusions about the effect of living in a high poverty neighborhood during young adulthood on mid-to-late life health. The results reveal that even a large amount of selection on unobservable factors does not eliminate the significant effect of neighborhood poverty on health status later in life. This evidence indicates further research on the effects of particular neighborhood characteristics is strongly warranted to identify the causal mechanisms through which concentrated neighborhood poverty effects operate.

The findings in this paper further our understanding of the underlying processes that produce health disparities between different racial and socioeconomic groups. The study finds that racial differences in later-life health can be largely accounted for by childhood socioeconomic status and young adult family and neighborhood factors, while contemporaneous neighborhood factors account for relatively little of this gap. Previous research has shown health behaviors, such as cigarette smoking, alcohol use, diet and exercise, explain a relatively small amount of race and socioeconomic differences in health status (see e.g., Lantz et al., 2001). The implications of these prior findings is that short-term public health policies focused on individual-level risk factors such as smoking, alcohol consumption, obesity, and exercise, are not likely to be an effective means of addressing race and SES differences in health. This paper instead highlights the important role of differences in family and neighborhood socioeconomic conditions experienced earlier in life in contributing to racial health disparities in mid to late life. The paper investigates differential effects of socioeconomic conditions on health over the life cycle and provides evidence that dimensions of neighborhood disadvantage that occur earlier in life are more predictive of morbidity during mid-to-late life than are contemporaneous neighborhood measures. The cumulative and dynamic nature of the relationship between socioeconomic conditions and health over the life course applies to dimensions of both family and neighborhood characteristics. The findings of this paper challenge future research to further our understanding of the pathways through which neighborhoods influence health and how the effects of neighborhood conditions differ over the life cycle. This knowledge is critical to help policy makers develop interventions (e.g., early childhood interventions or targeted policies for the geographic deconcentration of the poor) that build a bridge between childhood and early adulthood for impoverished families, so fewer individuals arrive at the doorstep of retirement with accumulated exposures that are irreversible.

---

[1] Differentials in the likelihood of reaching old age are important in understanding the health differences observed in old age (Johnson, 2008). Selective mortality is a potentially important issue for analyses of later-life health, and we investigated strategies to eliminate the potential for induced selection biases in our estimates, which could lead to an underestimate of the role of neighborhoods if ignored. (A summary of sample attrition is discussed in the Appendix).

[2] The 1968 addresses were geocoded to census block identifiers using GDT geographic mapping technologies. Census blocks are the smallest level of geographic precision reported by the Census Bureau and represent a narrow definition of neighborhood. Census block identifiers are defined for the entire U.S. in 2000.

[3] A contribution of this work pertains to measurement and the conceptualization of neighborhood contexts. We analyze the sensitivity of the main results to the level of aggregation used, and compare how the results differ if we instead use larger neighborhood constructs for neighborhood groupings (e.g., census block vs. census tract vs. zip code or county).

[4] We also analyzed the onset of health-limiting conditions and found a similar pattern of results as reported in the paper (results available from authors upon request).

[5] The key shortcoming of an ordered logit or ordered probit regression is the probit and logit link functions are inadequate to model health due to the significant degree of skewness in the health distribution (i.e., the majority of a general population sample report themselves to be in good to excellent health). Van Doorslaer and Jones (2003) assess the validity of using ordered probit regressions to impose cardinality on the ordinal responses comparing it with a gold standard of using the McMaster 'Health Utility Index Mark III' (HUI). They conclude "…the ordered probit regression does not allow for any sensible approximation of the true degree of inequality."

[6] To be eligible for the SEO sample, households had to have income that was below two times the poverty line, which in theory could be problematic for our purposes because two neighboring families could enter that component of the PSID only if they had sufficiently low income. However, due to the significant degree of residential segregation by income, we find evidence that the typical neighbor of a low-income family was also low income; thus, in practice this does not present any significant within-neighborhood sample selection bias problems. In particular, in the 1968 SRC component of the PSID, the average family with income less than two times the poverty line (in that year) lived in neighborhoods in which neighbors' average income was also among the bottom third of the income distribution. Similarly, using larger national samples geocoded to the census block, Hardman and Ioannides (2005) find that among the poorest 30 percent of households, roughly 75 percent live in neighborhoods in which neighbors' median income is also among the poorest 30 percent of households.

[7] These measures serve as proxies of neighborhood quality as this information was only collected in the 1975 survey and may not reflect the characteristics of the 1968 neighborhood due to residential mobility over the period. However, 1968 families in the PSID tended to move to neighborhoods that had observable neighborhood characteristics that were similar to their previous residential location (Kunz et al, 2001).

[8] Raudenbush, Sampson, and Morenoff have pioneered these approaches of surveying multiple respondents from the same neighborhood for the measurement of neighborhood characteristics in their Project on Human Development in Chicago Neighborhoods (PHDCN) study.

[9] This discussion follows Solon et al (2000).

[10] Maximum-likelihood (ML) estimates based on a numerical integration procedure were computed using aML statistical software (Lillard and Panis, 2003). Some estimates were computed using the gllamm macro in Stata (Rabe-Hesketh et al, 2000). The numerical evaluation of the unconditional-likelihood function uses Gaussian quadrature. We use 10-point quadrature for each level.

[11] Assortative mating patterns may result in similarity of hereditary risk factors between spouses.

[12] We examined alternative functional forms of the key explanatory variables to best fit the data.

[13] This magnitude may be understated due to downward errors-in-variables bias from retrospective reports of childhood SES.

Appendix

PSID sample

The sample consists of PSID respondents who were young adults when the study began and who have been followed into mid-to-late adulthood; they were born between 1928 and 1948 and were between 20 and 39 years old in 1968. We obtain all available information on them for each wave, 1968 to 2005. In 2005, the oldest respondent is 77 and the youngest is 57.

The first wave of PSID interviewing in 1968 included 4,354 individuals who were in their 20s and 30s. 321 of these individuals died by 2005. These individuals are included in the analyses for the years they are observed alive. Johnson (2009) analyzes the effects of family and neighborhood conditions earlier in life on adult mortality in mid-to-late life. The results suggest that any selective attrition with respect to mortality is likely to lead to an understatement of the impact of adverse childhood conditions because those who suffered premature death disproportionately grew up in the more disadvantaged childhood family and neighborhood environments. Of the 4,354 young adults present in the initial wave of the PSID, 2,871 had at least one valid report of health status in adulthood. Adult GHS is based on reports for PSID heads and wives/"wives" (1984-2005) as well as all family members in 1986. A small minority of respondents lacked valid addresses and were not able to be matched to neighborhoods in the geocode file—these cases were disproportionately located in rural areas. The resultant sample used in the analyses contains 2,730 individuals that came from 1,894 different families, 1,457 neighborhoods, and 306 counties. Data are combined across all waves for each person, and in total there are 36,121 person-year observations, or an average of 13 observations per person, for the analyses of adult health.

While the decline in the initial sample of 37 percent is substantial, it is low given the long period over which these children and their families are followed. For example, among the 17,287 newborns participating in the 1970 British birth cohort sample, 6,454 (37 percent) were not interviewed (i.e., were not in the "observed sample") in 1999/2000 when they were 30 years old. Moreover, studies have concluded that the PSID sample of heads and wives remains representative of the national sample of adults (Fitzgerald, Gottschalk, and Moffitt, 1998a; Becketti et al, 1988), and that the sample of "split offs" is representative (Fitzgerald, Gottschalk and Moffitt, 1998b). The 95-98% wave-to-wave response rate of the PSID makes this possible.

Table A0 contains a summary of the variable definitions and data sources of all key measures used in the analyses, the year(s) of data collection, and the relevant survey questions used to construct these measures. Table A1 reports descriptive statistics for the samples used in the models of adult health status both for the full sample and separately by race. The substantial race differences in childhood and young adult family and neighborhood characteristics are highlighted in this table.

Health Index

A number of previous studies using surveys have demonstrated that a change in GHS from fair to poor represents a much larger degree of health deterioration than a change from excellent to very good or very good to good (e.g., Van Doorslaer and Jones, 2003; Humphries and Van Doorslaer, 2000). More generally, this research has shown that health differences between GHS categories are larger at lower levels of GHS. Thus, assuming a linear scaling would not be appropriate.

To analyze health disparities in the presence of a multiple-category health indicator, three alternative approaches have been used, each with its own set of advantages and disadvantages. The most common and simplest approach is to dichotomize GHS by setting a cut-off point above which individuals are said to be in good health (e.g., excellent/very good/good vs. fair/poor). The disadvantage of this approach is that it does not utilize all of the information on health. Additionally, it uses a somewhat arbitrary cut-off for the determination of healthy/not-healthy, and the measurement of inequality over time can be sensitive to the choice of cut-off (Wagstaff and Van Doorslaer, 1994).

A second approach is to estimate an ordered logit or ordered probit regression using the GHS categories as the dependent variable, and rescale the predicted underlying latent variable of this model to compute "quality weights" for health between 0 and 1 (Cutler and Richardson, 1997; Groot, 2000). The key shortcoming of this approach is the probit and logit link functions are inadequate to model health due to the significant degree of skewness in the health distribution (i.e., the majority of a general population sample report themselves to be in good to excellent health). Van Doorslaer and Jones (2003) assess the validity of using ordered probit regressions to impose cardinality on the ordinal responses comparing it with a gold standard of using the McMaster 'Health Utility Index Mark III' (HUI).[1] They conclude "…the ordered probit regression does not allow for any sensible approximation of the true degree of inequality."

The third approach, adopted first by Wagstaff and Van Doorslaer (1994), assumes that underlying the categorical empirical distribution of the responses to the GHS question is a latent, continuous but unobservable health variable with a standard lognormal distribution. This assumption allows "scoring" of the GHS categories using the mid-points of the intervals corresponding to the standard lognormal distribution. The lognormal distribution allows for skewness in the underlying distribution of health. The health inequality results obtained using this scaling procedure have been shown to be comparable to those obtained using truly continuous generic measures like the SF36 (Gerdtham et al., 1999) or the Health Utility Index Mark III (Humphries and van Doorslaer, 2000) in Canada, but has not been validated as an

[1] The McMaster Health Utility Index can be considered a more objective health measure because the respondents are only asked to classify themselves into eight health dimensions: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain. The Health Utility Index Mark III is capable of describing 972,000 unique health states (Humphries and van Doorslaer, 2000).

appropriate scaling procedure using U.S. data. The disadvantage of this approach is it inappropriately uses OLS on what remains essentially a categorical variable and does not exploit the within-category variation in health. This is particularly problematic for the analysis of health dynamics over a relatively short time horizon. Ignoring within-category variation in health will cause health deterioration estimates to be biased and induce (health) state dependence because within-category variation increases when going down from excellent to poor health.

Several surveys have been undertaken that contain both the GHS question and questions underlying a health utility index. In this paper, we adopt a latent variable approach that combines the advantages of approaches two and three above, but avoids their respective pitfalls. Specifically, utilizing external U.S. data that contain both GHS and health utility index measures, we use the distribution of health utility-based scores across the GHS categories to scale the categorical responses and subject our indicators to the transformation that best predicts quality of life. This scaling thus translates our measures into the metric that reflects the underlying level of health. Specifically, using a 100-point scale where 100 equals perfect health and zero is equivalent to death, the interval health values associated with GHS are: [95, 100] for excellent, [85, 95) for very good, [70,85) for good, [30,70) for fair, and [1,30) for poor health.

Interval Regression Model. The method assumes that underlying the categorical empirical distribution of the responses to the GHS question is a latent, continuous health variable. We estimate interval regression models using the aforementioned values to scale the thresholds for GHS, where interval regression models are equivalent to probit models with known thresholds.

The measure of health status has categorical outcomes excellent (E), very good (VG), good (G), fair (F), and poor (P). The model can be expressed as

$$H_i = \begin{cases} 1 \text{ (E)} & \text{if } 95 \le H_i^* \le 100 = \text{perfect health} \\ 2 \text{ (VG)} & \text{if } 85 \le H_i^* < 95 \\ 3 \text{ (G)} & \text{if } 70 \le H_i^* < 85 \\ 4 \text{ (F)} & \text{if } 30 \le H_i^* < 70 \\ 5 \text{ (P)} & \text{if } 1 \le H_i^* < 30, \end{cases}$$

where H^* is the continuous latent health variable and is assumed to be a function of socio-economic variables x:

$$H_i^* = x_i\beta + v_i, \quad v_i \sim N(0, \sigma_v^2).$$

Given the assumption that the error term is normally distributed, the probability of observing a particular value of y is

$$P_{ij} = P(H_i = j) = \Phi\left(\frac{\mu_U - x_i\beta}{\sigma_v}\right) - \Phi\left(\frac{\mu_L - x_i\beta}{\sigma_v}\right),$$

where j indexes the categories, $\Phi(\bullet)$ is the standard normal distribution function, and μ represent the threshold values previously discussed. Because the threshold values are known, it is possible to identify the variance of the error term σ_v^2. Because we use the health utility-based values to score the thresholds for GHS, the linear index for the interval regression model is measured on the same scale. This scaling thus translates the measures into the metric that reflects the underlying level of health. With independent observations, the log-likelihood for the interval regression model takes the form:

$$\log L = \sum_i \sum_j H_{ij} \log P_{ij} \quad ,$$

where the H_{ij} are binary variables that are equal to 1 if $H_{ij} = j$. This can be maximized to give estimates of β.

References

Aaronson, Daniel. 1998. "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes". *The Journal of Human Resources*. Autumn, 33(4): 915-946.

Altonji, Joseph, Todd Elder and Christopher Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, Vol. 113(1):151-184.

Blackwell, Debra, M. Hayward, and E. Crimmins. 2001. "Does Childhood Health Affect Chronic Morbidity in Later Life?" *Social Science and Medicine*. 52:1269-84.

Case, A., Fertig, A., and C. Paxson. (2005). "The Lasting Impact of Childhood Health and Circumstances." *Journal of Health Economics*, 24: 365-389.

Case Anne and L. Katz. 1991. "The Company you Keep: the Effects of Family and Neighborhood on Disadvantaged Families". National Bureau of Economic Research Working Paper no. 3705.

Case, Anne and Angus Deaton. 2003. "Broken Down by Work and Sex: How our Health Declines". NBER Working Paper. Center for Health and Wellbeing, Princeton University.

Case, Anne, Darren Lubotsky and Christian Paxson. 2002. "Economic Status and Health in Childhood: The Origins of the Gradient". *American Economic Review*. December, 92(5): 1308-1334.

Chandra, Amitabh and Jonathan S. Skinner. 2003. "Geography and Racial Health Disparities". Unpublished manuscript. Department of Economics and Center of Evaluative Clinical Sciences, Dartmouth Medical School.

Corcoran, M., R. Gordon, D. Laren, and G. Solon. 1992. "The Association Between Men's Economic Status and their Family and Community Origins". *Journal of Human Resources*. 27: 575-601.

Corcoran, M., C. Jencks, and M. Olneck. 1976. "Intergenerational Transfers of Inequality: The Effects of Family Background on Earnings". *American Economic Review*. 66(2):430-35.

Crane J. 1991. "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping out and Teenage Childbearing". *American Journal of Sociology.* 96: 1226-1259.

Currie, Janet and Rosemary Hyson. 1999. "Is the Impact of Health Shocks Cushioned by Socioeconomic Status? The Case of Low Birthweight". *The American Economic Review*. May, 89(2): 245-250.

Cutler, David M. and Edward L. Glaeser (1997). "Are Ghettos Good or Bad?" *Quarterly Journal of Economics* 112: 827-872.

Cutler, David, E. Richardson, and T. Keeler. 1997. "Measuring the Health of the US Population". *Brookings Papers on Economic Activity. Microeconomics*. 1997:217-282.

Deaton, Angus and Darren Lubotsky, 2002. "Mortality, Inequality and Race in American Cities and States." *Social Science and Medicine* 56(6): 1139-1153.

Deaton, Angus and Christina Paxson. 1998. "Health, Income, and Inequality over the Life Cycle". In *Frontiers in the Economics of Aging*, David Wise, Ed. Chicago: University of Chicago Press. 431-462.

Duncan, Greg, J. Boisjoly, and K. M. Harris. 2001. "Sibling, Peer, Neighbor, and Schoolmate Correlations as Indiciators of Importance of Context for Adolescent Development". *Demography*. August 38 (3):437-447.

Ellen, IG, T. Mijanovich, and K. Dillman. 2001. “Neighborhood effects on health: exploring the links and assessing the evidence”. *Journal of Urban Affairs.* 23(3): 391-408.

Evans W.N., W. Oates and R.M. Schwab. 1992. “Measuring peer group effects: a study of teenage behavior”. *Journal of Political Economy.* 100:966-91.

Fan, Jianqing. 1992. “Design-Adaptive Nonparametric Regression”. *Journal of the American Statistical Association*. 87:998-1004.

Hart, Kevin D., Stephen J. Kunitz, Ralph R. Sell, and Dana B. Mukamel (1998). “Metropolitan Governance, Residential Segregation, and Mortality among African Americans.” *American Journal of Public Health* 88(3): 434-438.

Hayward, Mark D. and Bridget K. Gorman. “The Long Arm of Childhood: The Influence of Early Life Social Conditions on Men’s Mortality”. Unpublished manuscript. Population Research Institute, Penn State University.

Heckman, James. 2007. Economics of Health and Mortality Special Feature: “The Economics, Technology, and Neuroscience of Human Capability Formation”. *Proceedings of the National Academy of Sciences* 104:13250–13255.

Hill, Martha S. 1992. The Panel Study of Income Dynamics: A User's Guide to Major Social Science Data Bases, Vol. 2. Newbury Park, CA: Sage Publications, Inc.

Johnson, Rucker C. 2008. “Health Dynamics of the Evolution of Health Inequality over the Life Course: The Importance of Neighborhood & Family Background”. Unpublished manuscript, UC-Berkeley.

_______. 2008. “Inequality in Men’s Mortality: The Influence of Young Adult Neighborhood and Family Socioeconomic Factors”. Unpublished manuscript, UC-Berkeley.

Johnson, Rucker C. and Robert F. Schoeni. 2007. “The Influence of Early-Life Events on Human Capital, Health Status, and Labor Market Outcomes Over the Life Course”. National Poverty Center working paper.

Katz, L., J. Kling and J. Liebman. 2001. “Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment”. *Quarterly Journal of Economics*. 116(2):607-654.

Kawachi, I. and L. Berkman (2003). *Neighborhoods and Health*. Oxford University Press, USA.

Krauth, Brian. 2006. "Simulation-based estimation of peer effects". *Journal of Econometrics* 133(1): 243-271.

Kremer, Michael. 1997. “How Much Does Sorting Increase Inequality?” *Quarterly Journal of Economics*. Feb., 112(1): 115-139.

Kunz, Jim, Marianne E. Page and Gary Solon. 2001. “Are Single-Year Measures of Neighborhood Characteristics Useful Proxies for Children’s Long-Run Neighborhood Environment?” Unpublished manuscript. University of Maryland and Joint Center for Poverty Research.

Leventhal, T and J. Brooks-Gunn. 2001. “Moving to Opportunity: What About the Kids?” forthcoming in *Choosing a Better Life: Evaluating the Moving to Opportunity Social Experiment*, eds. J. Goering and J. Ferris. Washington, DC: Urban Institute Press.

Manski, Charles F. 1993. “Identification of Endogenous Social Effects: The Reflection Problem”. *The Review of Economic Studies*. 60(3): 531-542.

Moffitt, R. 1998. “Policy Interventions, Low-Level Equilibria and Social Interactions”. Department of Economics, Johns Hopkins University.
Morenoff, Jeff, and John Lynch. 2002. “What makes a place healthy? Neighborhood influences on racial/ethnic disparities in health over the life course.” Paper prepared for a Workshop on Ethnic Disparities in Aging Health, Washington, DC.
Murray, C.J.L., Michaud, C.M., McKenna, M.T., and Marks, J.S. (1998). U.S. patterns of mortality by county and race: 1965-1994. *Harvard Center for Population and Development Studies* (Monograph Series). Cambridge, MA.
Oreopoulos, Phillip. 2002. “The Long-Run Consequences of Living in a Poor Neighborhood”. Unpublished manuscript.
Page, Marianne E. and Gary Solon. 1999. “Correlations between Brothers and Neighboring Boys in their Adult Earnings: The Importance of Being Urban”. Unpublished manuscript. University of California at Davis.
_______. 2001. “Correlations between Sisters and Neighboring Girls in Their Subsequent Income as Adults”. Unpublished manuscript. University of California at Davis.
Plotnick, R. and S. Hoffman. 1996. “The Effect of Neighborhood Characteristics on Young Adult Outcomes: Alternative Estimates”. Institute for Research on Poverty, Discussion Paper 1106-96.
Raudenbush, S. and S.W. Bryk. 2002. *Hierarchical Linear Models. Applications and Data Analysis Methods.* Thousand Oaks, CA: Sage Publications.
Roberts E.M. 1999. “Socioeconomic position and health: The independent contribution of community socioeconomic context”. *Annual Review of Sociology*. 25:489-516.
Smith, J.P. 2007. “The Impact of SES on Health Over the Life-Course.” *Journal of Human Resources*.
Smith, J.P. 1999. “Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Economic Status,” *Journal of Economic Perspectives* 13(2):, Spring 1999.
Solon, Gary, M. Corcoran, R. Gordon, and D. Laren. 1987. “The Effect of Family Background on Economic Status: a Longitudinal Analysis of Sibling Correlations”. NBER Working Paper Series. Paper #2282. National Bureau of Economic Research.
Solon, Gary, Marianne Page, and Greg Duncan. 2000. “Correlations between neighboring children in their subsequent educational Attainment”. *The Review of Economics and Statistics*. 82(3): 383-392.
Tiebout, Charles M. 1956. “A Pure Theory of Local Expenditures”. *Journal of Political Economy*. 64: 416-424.
Yen I.H. and S. L. Syme. 1999. “The social environment and health: A discussion of the epidemiologic literature”. *Annual Review of Public Health*. 20:287-308.

Figure 1.

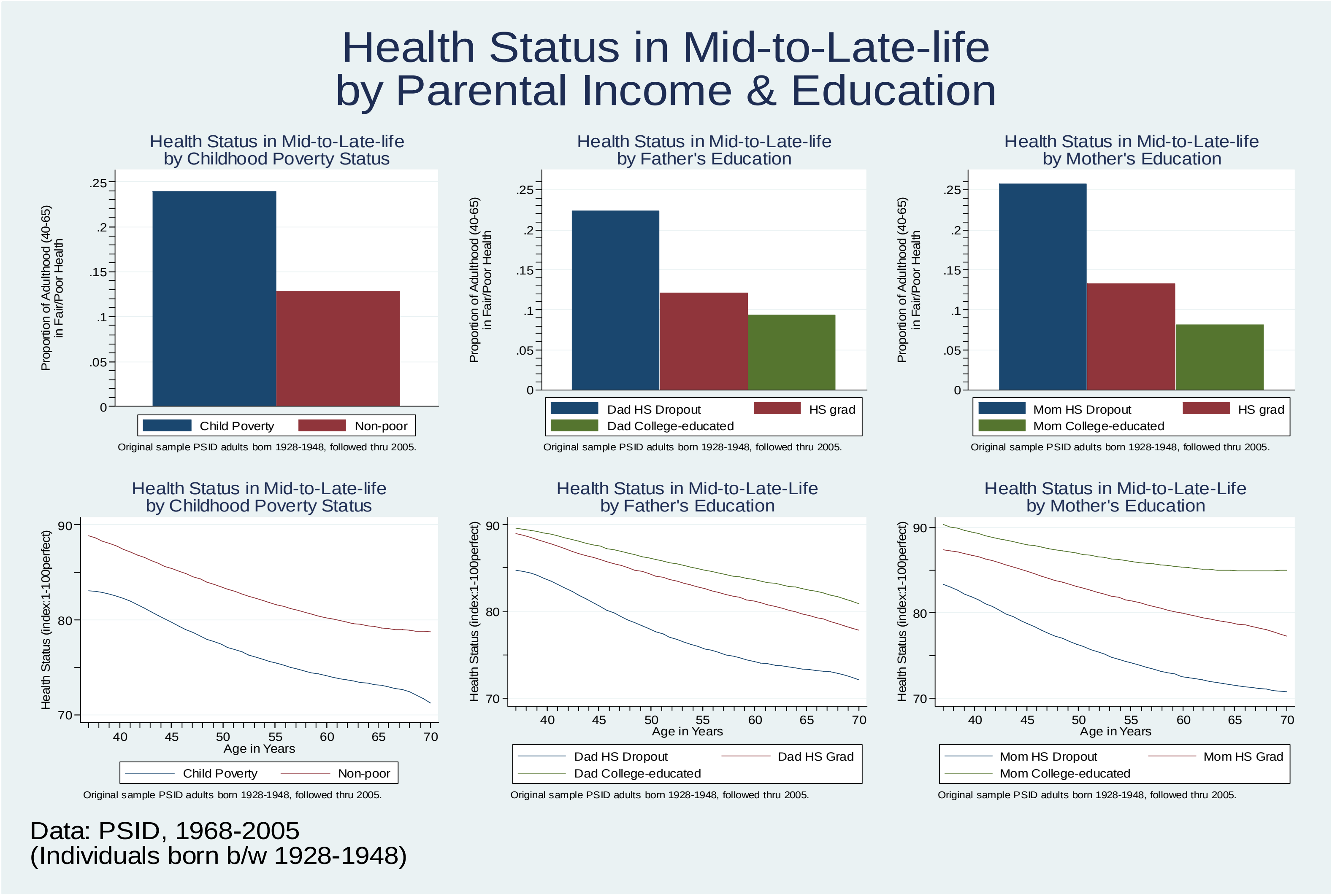

Health Status in Mid-to-Late-life
by Parental Income & Education
Health Status in Mid-to-Late-life
by Childhood Poverty Status
Proportion of Adulthood (40-65)
in Fair/Poor Health
Child Poverty
Non-poor
Original sample PSID adults born 1928-1948, followed thru 2005.
Health Status in Mid-to-Late-life
by Father's Education
Dad HS Dropout
HS grad
Dad College-educated
Original sample PSID adults born 1928-1948, followed thru 2005.
Health Status in Mid-to-Late-life
by Mother's Education
Mom HS Dropout
HS grad
Mom College-educated
Original sample PSID adults born 1928-1948, followed thru 2005.
Health Status in Mid-to-Late-life
by Childhood Poverty Status
Health Status (index:1-100perfect)
Age in Years
Child Poverty
Non-poor
Original sample PSID adults born 1928-1948, followed thru 2005.
Health Status in Mid-to-Late-Life
by Father's Education
Dad HS Dropout
Dad HS Grad
Dad College-educated
Original sample PSID adults born 1928-1948, followed thru 2005.
Health Status in Mid-to-Late-Life
by Mother's Education
Mom HS Dropout
Mom HS Grad
Mom College-educated
Original sample PSID adults born 1928-1948, followed thru 2005.
Data: PSID, 1968-2005
(Individuals born b/w 1928-1948)

Figure 2.

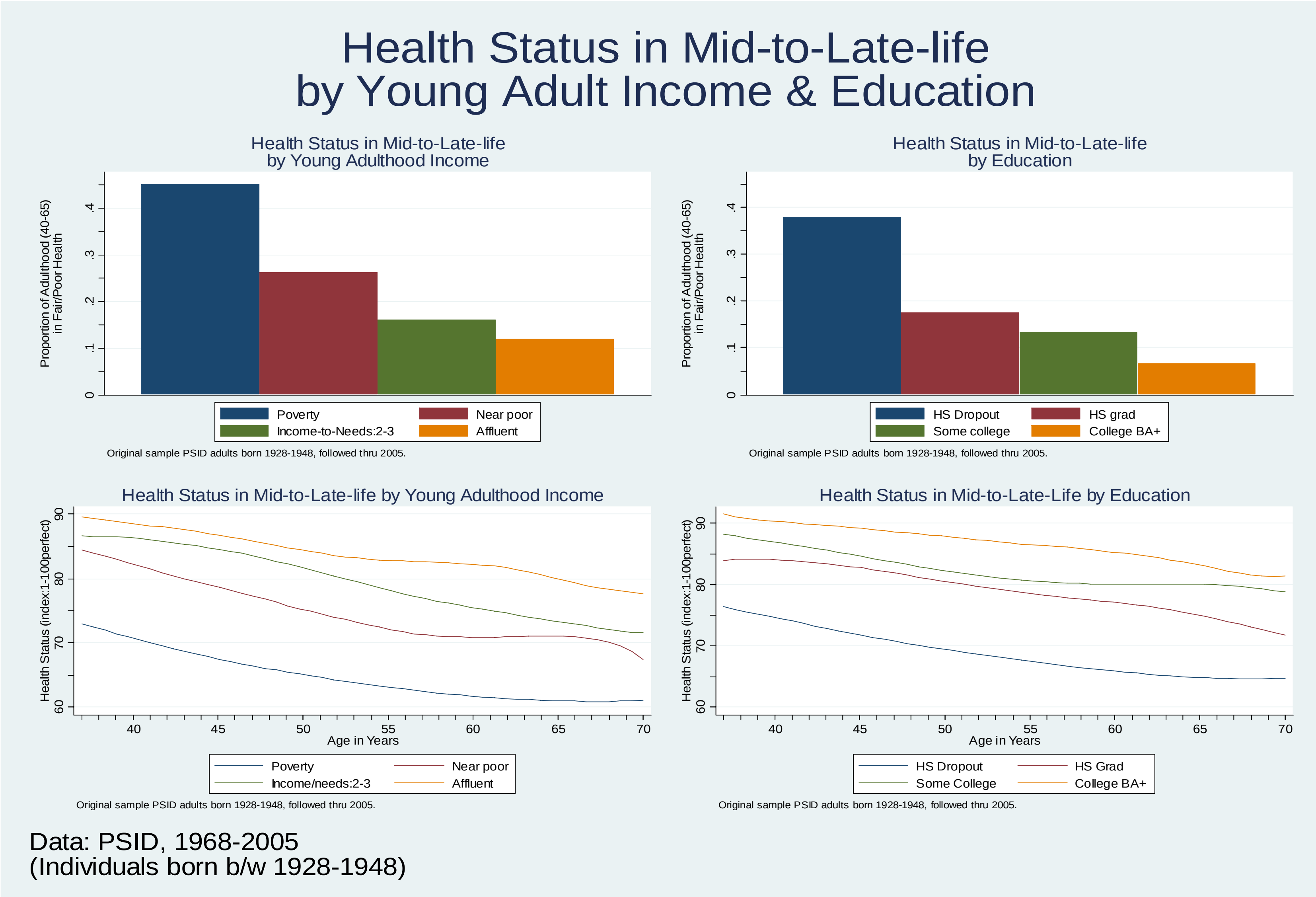

Health Status in Mid-to-Late-life
by Young Adult Income & Education
Health Status in Mid-to-Late-life
by Young Adulthood Income
Proportion of Adulthood (40-65)
in Fair/Poor Health
0
.1
.2
.3
.4
Poverty
Near poor
Income-to-Needs:2-3
Affluent
Original sample PSID adults born 1928-1948, followed thru 2005.
Health Status in Mid-to-Late-life
by Education
Proportion of Adulthood (40-65)
in Fair/Poor Health
0
.1
.2
.3
.4
HS Dropout
HS grad
Some college
College BA+
Original sample PSID adults born 1928-1948, followed thru 2005.
Health Status in Mid-to-Late-life by Young Adulthood Income
Health Status (index:1-100perfect)
60
70
80
90
40
45
50
55
60
65
70
Age in Years
Poverty
Near poor
Income/needs:2-3
Affluent
Original sample PSID adults born 1928-1948, followed thru 2005.
Health Status in Mid-to-Late-Life by Education
Health Status (index:1-100perfect)
60
70
80
90
40
45
50
55
60
65
70
Age in Years
HS Dropout
HS Grad
Some College
College BA+
Original sample PSID adults born 1928-1948, followed thru 2005.
Data: PSID, 1968-2005
(Individuals born b/w 1928-1948)

Figure 3.

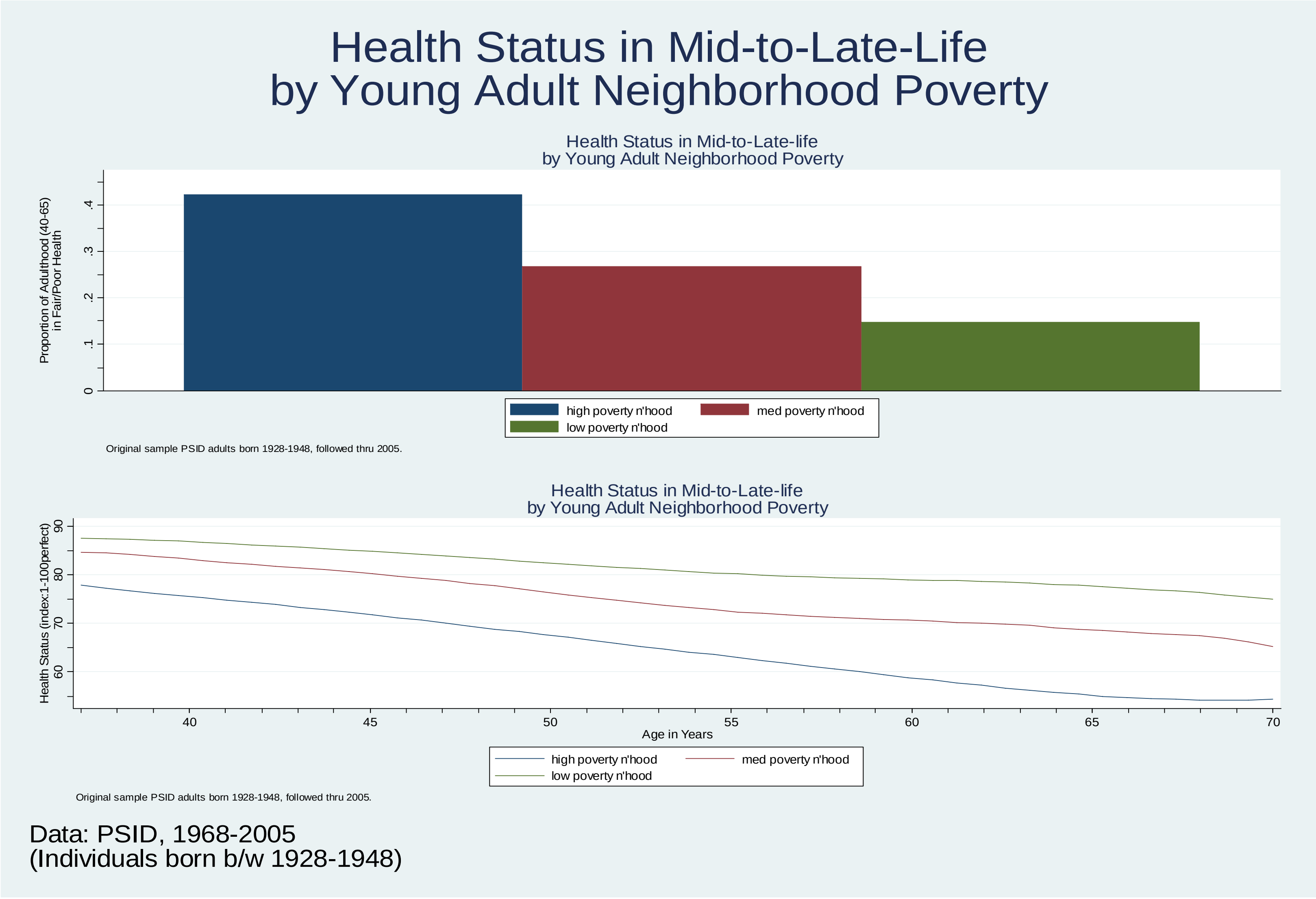
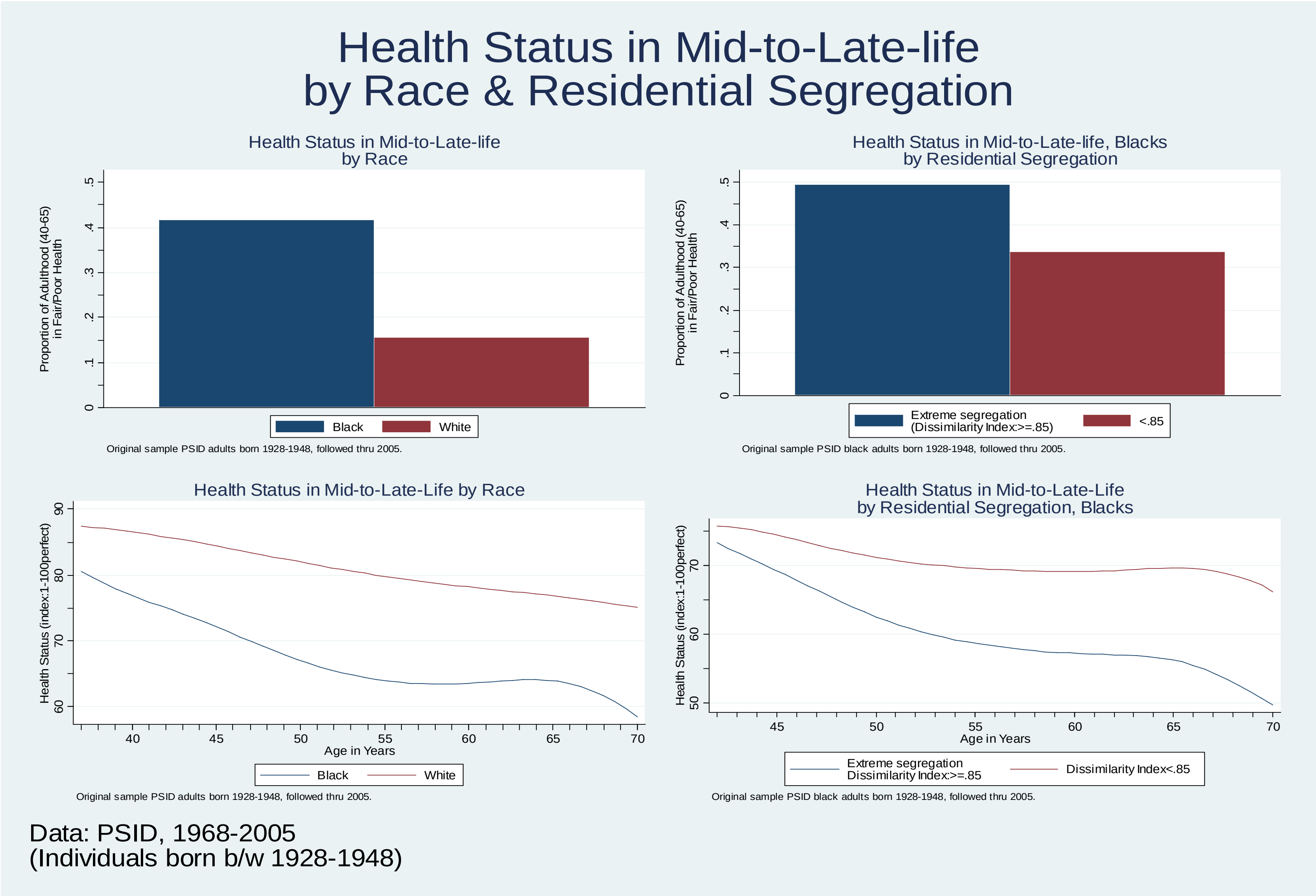

Health Status in Mid-to-Late-Life
by Young Adult Neighborhood Poverty
Health Status in Mid-to-Late-life
by Young Adult Neighborhood Poverty
Proportion of Adulthood (40-65) in Fair/Poor Health
0
.1
.2
.3
.4
high poverty n'hood
med poverty n'hood
low poverty n'hood
Original sample PSID adults born 1928-1948, followed thru 2005.
Health Status in Mid-to-Late-life
by Young Adult Neighborhood Poverty
Health Status (index:1-100perfect)
60
70
80
90
40
45
50
55
60
65
70
Age in Years
high poverty n'hood
med poverty n'hood
low poverty n'hood
Original sample PSID adults born 1928-1948, followed thru 2005.
Data: PSID, 1968-2005
(Individuals born b/w 1928-1948)

Figure 4.



Health Status in Mid-to-Late-life
by Race & Residential Segregation
Health Status in Mid-to-Late-life
by Race
Proportion of Adulthood (40-65)
in Fair/Poor Health
0
.1
.2
.3
.4
.5
Black
White
Original sample PSID adults born 1928-1948, followed thru 2005.
Health Status in Mid-to-Late-life, Blacks
by Residential Segregation
Proportion of Adulthood (40-65)
in Fair/Poor Health
0
.1
.2
.3
.4
.5
Extreme segregation
(Dissimilarity Index:>=.85)
<.85
Original sample PSID black adults born 1928-1948, followed thru 2005.
Health Status in Mid-to-Late-Life by Race
Health Status (index:1-100perfect)
60
70
80
90
40
45
50
55
60
65
70
Age in Years
Black
White
Original sample PSID adults born 1928-1948, followed thru 2005.
Health Status in Mid-to-Late-Life
by Residential Segregation, Blacks
Health Status (index:1-100perfect)
50
60
70
45
50
55
60
65
70
Age in Years
Extreme segregation
Dissimilarity Index:>=.85
Dissimilarity Index<.85
Original sample PSID black adults born 1928-1948, followed thru 2005.
Data: PSID, 1968-2005
(Individuals born b/w 1928-1948)

Figure 5.

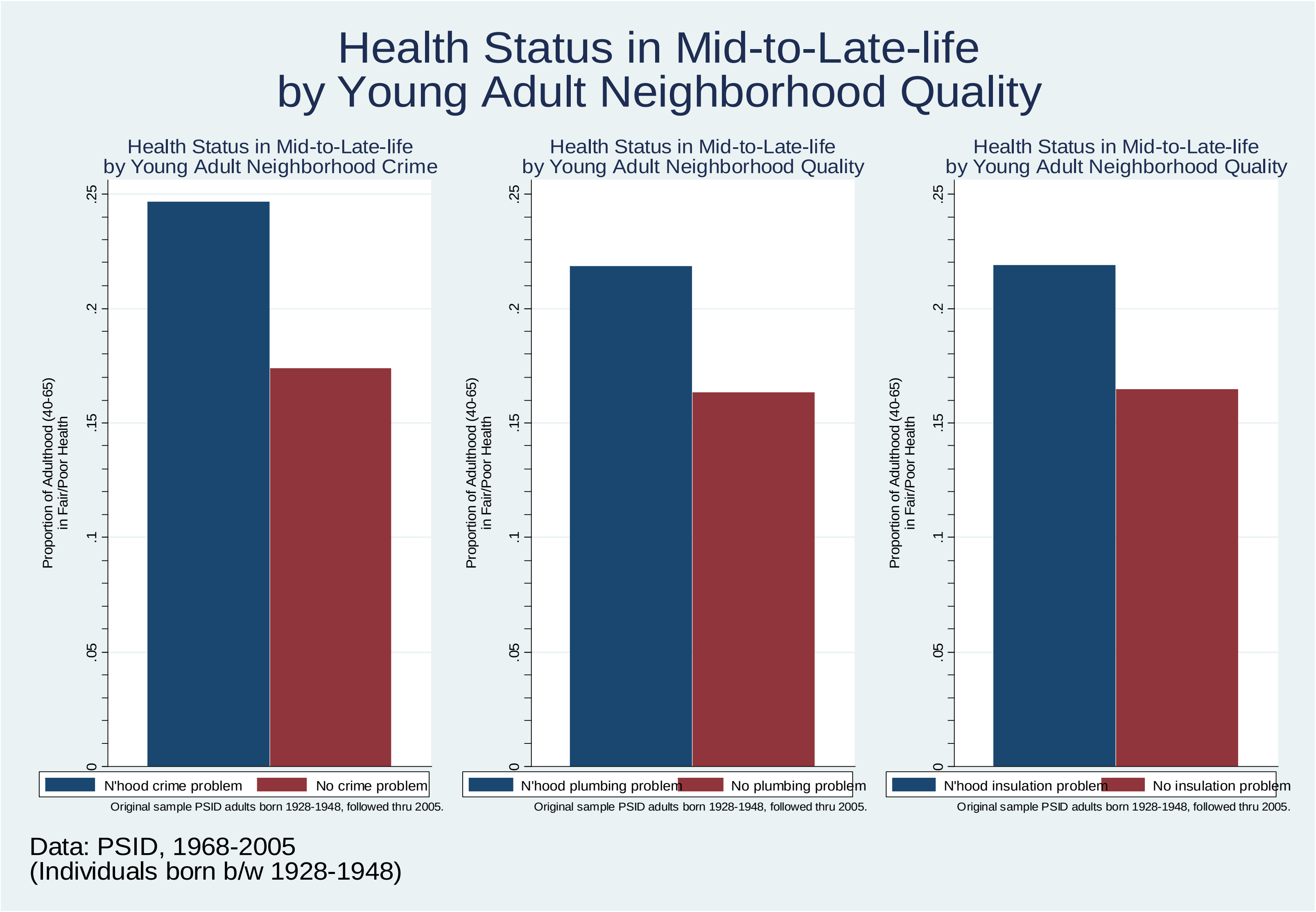

Health Status in Mid-to-Late-life
by Young Adult Neighborhood Quality
Health Status in Mid-to-Late-life
by Young Adult Neighborhood Crime
Health Status in Mid-to-Late-life
by Young Adult Neighborhood Quality
Health Status in Mid-to-Late-life
by Young Adult Neighborhood Quality
Proportion of Adulthood (40-65)
in Fair/Poor Health
0
.05
.1
.15
.2
.25
N'hood crime problem
No crime problem
N'hood plumbing problem
No plumbing problem
N'hood insulation problem
No insulation problem
Original sample PSID adults born 1928-1948, followed thru 2005.
Data: PSID, 1968-2005
(Individuals born b/w 1928-1948)

Table 1. Mid to Late-life Health: Importance of Young Adult Neighborhood & Family Background

(Dependent variable: general health status) Hierarchical Random Effects Interval Regression Model: 100pt-scale, 100=perfect health			
	All Adulthood yrs 35+	Ages 35-55	Over 55
	(1)	(2)	(3)
Constant	81.0112***	81.1520***	77.9502***
	(0.1221)	(0.1224)	(0.1631)
Age - 50	-0.3909***	-0.4340***	
	(0.0026)	(0.0043)	
Age - 60			-0.4451***
			(0.0062)
Female	-1.8432***	-2.1336***	-1.7299***
	(0.1295)	(0.1293)	(0.1801)
Random Effects, Unmeasured (Std Dev)			
Young Adult Neighborhood component	9.4605***	9.5284***	10.5244***
	(0.1262)	(0.1343)	(0.1718)
Family component	5.8912***	5.6670***	5.0742***
	(0.1921)	(0.2079)	(0.3605)
Individual component	12.0808***	11.7314***	15.0163***
	(0.0727)	(0.0732)	(0.1070)
Transitory error component	7.9767***	6.7325***	7.7180***
	(0.0075)	(0.0082)	(0.0122)
Log-likelihood	-2235086	-1338205	-875160.66
Number of counties	306	306	306
Number of neighborhoods	1,457	1,416	1,129
Number of families	1,894	1,827	1,460
Number of individuals	2,730	2,625	2,019
Number of person-year observations	36,121	22,766	13,355

*** p<0.01, ** p<0.05, * p<0.10

Note: Robust standard errors in parentheses and all standard errors are Huber-corrected, clustered on county. All models control for year of birth and include age squared and age cubed terms (coefficients suppressed to conserve space).

Table 2. Race & SES Differences in Mid to Late-life Health (Age 35+): Importance of Young Adult Neighborhood & Family Background

(Dependent variable: general health status in adulthood)

4-Level Hierarchical Random Effects Interval Regression Model: 100pt-scale, 100=perfect healtl

	Uncond'l model	Raw race gap	Controls for Childhood SES	Controls for Childhood SES + Young Adult Family	Controls for Childhood SES + Young Adult Nhood + Fam	Controls for Childhood SES + Young Adult backgrd + Contemporaneous NHood
Childhood factors	(1)	(2)	(3)	(4)	(5)	(6)
Black		-13.9161***	-9.5431***	-5.6848***	-4.1673***	-4.3711***
Non-Hispanic white (reference category)		(0.3019)	(0.3050)	(0.2958)	(0.3152)	(0.3159)
Childhood poverty			-2.1446***	-1.6253***	-1.5271***	-1.5162***
Non-poor (reference category)			(0.1726)	(0.1660)	(0.1658)	(0.1660)
Mother's education:						
High school dropout			-3.7195***	-2.0822***	-2.0715***	-2.0721***
High school graduate (reference category)			(0.1695)	(0.1649)	(0.1656)	(0.1657)
College educated			2.5914***	1.5036***	1.7085***	1.6955***
			(0.2621)	(0.2526)	(0.2518)	(0.2521)
Father's education:						
High school dropout			-2.2438***	-1.6475***	-1.6006***	-1.5880***
High school graduate (reference category)			(0.1857)	(0.1787)	(0.1780)	(0.1782)
College educated			0.3646+	-0.4336+	-0.2658	-0.2422
			(0.2745)	(0.2652)	(0.2643)	(0.2646)
Father's occupation:						
White collar			0.3910+	-0.2402	-0.0596	-0.0841
Blue collar (reference category)			(0.2527)	(0.2442)	(0.2437)	(0.2439)
Grew up on farm			-2.0038***	-0.4882**	-0.6670***	-0.7117***
Grew up in large, urban MSA (ref. category)			(0.2263)	(0.2124)	(0.2161)	(0.2164)
Grew up in small town			-2.1844***	-1.7464***	-1.7592***	-1.7843***
			(0.1987)	(0.1851)	(0.1866)	(0.1869)
Grew up in different places			0.3709	-0.5787	0.1393	0.1549
			(0.7765)	(0.7284)	(0.7245)	(0.7252)
Young Adulthood factors						
Family income-to needs ratio (avg during 1967-1972), spline:						
Income-to-needs ratio*ratio is <1				11.2528***	11.0745***	11.1995***
				(1.1209)	(1.1232)	(1.1244)
Income-to-needs ratio* ratio is 1 to 3				2.5285***	1.8788***	1.9091***
				(0.1485)	(0.1498)	(0.1499)
Income-to-needs ratio* ratio is >3				0.1094*	0.0809	0.0946+
				(0.0665)	(0.0661)	(0.0662)
Educational attainment:						
High school dropout				-6.6644***	-6.2487***	-6.2924***
High school graduate (reference category)				(0.2109)	(0.2114)	(0.2116)
Some college				0.8778***	0.9435***	0.9513***
				(0.1950)	(0.1953)	(0.1955)
College graduate or higher				4.1242***	4.4527***	4.4596***
				(0.2031)	(0.2027)	(0.2029)
No Private HI coverage, 1968-1972				-2.6030***	-2.0974***	-2.1235***
				(0.1667)	(0.1670)	(0.1672)
Smoked cigarettes at some point, 1968-1972				-2.2780***	-2.1436***	-2.1618***
				(0.1710)	(0.1706)	(0.1708)
Annual alcohol expenditures (in $100's), 5-year average 1968-1972				0.0266**	0.0174+	0.0170+
				(0.0120)	(0.0121)	(0.0121)

Table 2 (cont'd). Race & SES Differences in Mid to Late-life Health (Age 35+): Importance of Young Adult Neighborhood & Family Background

(Dependent variable: general health status in adulthood)
4-Level Hierarchical Random Effects Interval Regression Model: 100pt-scale, 100=perfect health

	Uncond'l model	Raw race gap	Controls for Childhood SES	Controls for Childhood SES + Young Adult Family	Controls for Childhood SES + Young Adult Nhood + Fam	Controls for Childhood SES + Young Adult backgrd + Contemporaneous NHood
Young Adult Neighborhood factors	(1)	(2)	(3)	(4)	(5)	(6)
Neighborhood poverty rate (1970), spline:						
Low poverty neighborhood (reference category)						
Medium poverty neighborhood					-2.7756***	-2.9527***
					(0.2725)	(0.2730)
(Neighborhood poverty rate - 20)* rate is 10 to 30%					-1.9103***	-2.0820***
					(0.3743)	(0.3749)
High poverty neighborhood					-2.5821***	-2.6958***
					(0.4466)	(0.4472)
Neighborhood crime problem					-1.3322***	-1.3664***
					(0.2856)	(0.2860)
Residential segregation dissimilarity index, 1970 (MSA)					-4.5794***	-4.8788***
					(1.4345)	(1.4363)
Neighborhood housing quality index					-0.9396***	-0.9383***
					(0.0535)	(0.0535)
Neighborhood connectedness to informal sources of help					0.5783***	0.5728***
					(0.0583)	(0.0584)
Average Aspirations index in neighborhood					0.5318***	0.5322***
					(0.0527)	(0.0528)
Contemporaneous Neighborhood						
Neighborhood poverty rate						0.3891***
						(0.0268)
Random Effects, Unmeasured (Std Dev)						
Young Adult Neighborhood component	9.4605***	8.2464***	6.7203***	5.0222***	5.0184***	5.0256***
	(0.1262)	(0.1362)	(0.1480)	(0.1757)	(0.1577)	(0.1582)
Family component	5.8912***	6.1992***	6.0597***	5.6664***	5.3695***	5.3814***
	(0.1921)	(0.1836)	(0.1827)	(0.1960)	(0.1864)	(0.1866)
Individual component	12.0808***	12.0422***	12.0526***	11.9043***	11.9036***	11.9117***
	(0.0727)	(0.0719)	(0.0720)	(0.0721)	(0.0707)	(0.0708)
Transitory error component	7.9767***	7.9767***	7.9769***	7.9791***	7.9771***	7.9749***
	(0.0075)	(0.0075)	(0.0075)	(0.0075)	(0.0075)	(0.0075)
Log-likelihood	-2235086	-2234077.2	-2232563.3	-2230475.4	-2230015.9	-2229901.9
Number of counties	306	306	306	306	306	306
Number of neighborhoods	1,457	1,457	1,457	1,457	1,457	1,457
Number of families	1,894	1,894	1,894	1,894	1,894	1,894
Number of individuals	2,730	2,730	2,730	2,730	2,730	2,730
Number of person-year observations	36,121	36,121	36,121	36,121	36,121	36,121

*** p<0.01, ** p<0.05, * p<0.10

Note: All models include a constant and controls for age, age squared, age cubed, year of birth, gender, and columns (3)-(6) include controls for birth order, region of birth, and indices intended to capture long-term planning horizon (coefficients suppressed to conserve space).

Table 3. Race & SES Differences in Mid to Late-life Health (Age 35-55): Importance of Young Adult Neighborhood & Family Background

(Dependent variable: general health status in adulthood)
4-Level Hierarchical Random Effects Interval Regression Model: 100pt-scale, 100=perfect healtl

	Uncond'l model	Raw race gap	Controls for Childhood SES	Controls for Childhood SES + Young Adult Family	Controls for Childhood SES + Young Adult Nhood + Fam	Controls for Childhood SES + Young Adult backgrd + Contemporaneous NHood
Childhood factors	(1)	(2)	(3)	(4)	(5)	(6)
Black		-14.4949***	-10.2506***	-6.4238***	-5.2458***	-5.3266***
Non-Hispanic white (reference category)		(0.3040)	(0.3083)	(0.3035)	(0.3210)	(0.3218)
Childhood poverty			-1.3926***	-0.9583***	-0.8027***	-0.8010***
Non-poor (reference category)			(0.1725)	(0.1670)	(0.1662)	(0.1663)
Mother's education:						
High school dropout			-3.3589***	-1.8096***	-1.8141***	-1.8138***
High school graduate (reference category)			(0.1696)	(0.1660)	(0.1659)	(0.1660)
College educated			2.6136***	1.6477***	1.9316***	1.8945***
			(0.2584)	(0.2504)	(0.2487)	(0.2489)
Father's education:						
High school dropout			-2.6530***	-2.0027***	-1.9255***	-1.8857***
High school graduate (reference category)			(0.1844)	(0.1785)	(0.1773)	(0.1774)
College educated			0.0905	-0.5665**	-0.3117	-0.2771
			(0.2704)	(0.2626)	(0.2608)	(0.2611)
Father's occupation:						
White collar			0.8635***	0.2164	0.3893+	0.3786+
Blue collar (reference category)			(0.2513)	(0.2441)	(0.2427)	(0.2429)
Grew up on farm			-2.2893***	-0.9223***	-1.0149***	-1.1212***
Grew up in large, urban MSA (ref. category)			(0.2269)	(0.2153)	(0.2184)	(0.2187)
Grew up in small town			-2.1645***	-1.6966***	-1.6357***	-1.7062***
			(0.1994)	(0.1879)	(0.1885)	(0.1886)
Grew up in different places			2.6910***	1.3571*	2.1853***	2.1411***
			(0.7886)	(0.7486)	(0.7418)	(0.7421)
Young Adulthood factors						
Family income-to needs ratio (avg during 1967-1972), spline:						
Income-to-needs ratio*ratio is <1				12.6425***	12.0062***	12.0201***
				(1.1252)	(1.1222)	(1.1229)
Income-to-needs ratio* ratio is 1 to 3				2.2339***	1.6325***	1.6691***
				(0.1510)	(0.1515)	(0.1516)
Income-to-needs ratio* ratio is >3				0.1179*	0.0799	0.0822
				(0.0673)	(0.0666)	(0.0667)
Educational attainment:						
High school dropout				-6.4392***	-6.1404***	-6.1589***
High school graduate (reference category)				(0.2142)	(0.2136)	(0.2138)
Some college				0.4847**	0.6371***	0.6620***
				(0.1964)	(0.1957)	(0.1958)
College graduate or higher				3.7414***	4.0982***	4.1290***
				(0.2039)	(0.2026)	(0.2028)
No Private HI coverage, 1968-1972				-2.3005***	-1.7638***	-1.7383***
				(0.1696)	(0.1691)	(0.1692)
Smoked cigarettes at some point, 1968-1972				-2.4748***	-2.2742***	-2.2738***
				(0.1747)	(0.1737)	(0.1738)
Annual alcohol expenditures (in $100's), 5-year average 1968-1972				0.0185+	0.0052	0.0050
				(0.0122)	(0.0122)	(0.0122)

Table 3 (cont'd). Race & SES Differences in Mid to Late-life Health (Age 35-55): Importance of Young Adult Neighborhood & Family Background

(Dependent variable: general health status in adulthood)
4-Level Hierarchical Random Effects Interval Regression Model: 100pt-scale, 100=perfect healtl

	Uncond'l model	Raw race gap	Controls for Childhood SES	Controls for Childhood SES + Young Adult Family	Controls for Childhood SES + Young Adult Nhood + Fam	Controls for Childhood SES + Young Adult backgrd + Contemporaneous NHood
Young Adult Neighborhood factors	(1)	(2)	(3)	(4)	(5)	(6)
Neighborhood poverty rate (1970), spline:						
Low poverty neighborhood (reference category)						
Medium poverty neighborhood					-2.2588***	-2.3401***
					(0.2752)	(0.2759)
(Neighborhood poverty rate - 20)* rate is 10 to 30%					-1.1892***	-1.2353***
					(0.3765)	(0.3772)
High poverty neighborhood					-0.5325	-0.5253
					(0.4500)	(0.4507)
Neighborhood crime problem					-0.6165**	-0.6449**
					(0.2918)	(0.2920)
Residential segregation dissimilarity index, 1970 (MSA)					-2.0344+	-2.4281*
					(1.4509)	(1.4528)
Neighborhood housing quality index					-1.1363***	-1.1406***
					(0.0538)	(0.0538)
Neighborhood connectedness to informal sources of help					0.5998***	0.6015***
					(0.0589)	(0.0589)
Average Aspirations index in neighborhood					0.4689***	0.4689***
					(0.0530)	(0.0531)
Contemporaneous Neighborhood						
Neighborhood poverty rate						0.0714**
						(0.0322)
Random Effects, Unmeasured (Std Dev)						
Young Adult Neighborhood component	9.5284***	8.1560***	6.7804***	5.3427***	5.1491***	5.1865***
	(0.1343)	(0.1507)	(0.1656)	(0.1794)	(0.1766)	(0.1760)
Family component	5.6670***	6.1183***	5.9368***	5.6989***	5.4752***	5.4415***
	(0.2079)	(0.1976)	(0.1993)	(0.1948)	(0.1954)	(0.1972)
Individual component	11.7314***	11.6713***	11.6627***	11.5072***	11.4734***	11.4871***
	(0.0732)	(0.0722)	(0.0723)	(0.0711)	(0.0705)	(0.0706)
Transitory error component	6.7325***	6.7327***	6.7330***	6.7333***	6.7333***	6.7304***
	(0.0082)	(0.0082)	(0.0723)	(0.0082)	(0.0082)	(0.0082)
Log-likelihood	-1338205	-1337130.8	-1335681	-1333818.6	-1333338.8	-1333230.3
Number of counties	306	306	306	306	306	306
Number of neighborhoods	1,416	1,416	1,416	1,416	1,416	1,416
Number of families	1,827	1,827	1,827	1,827	1,827	1,827
Number of individuals	2,625	2,625	2,625	2,625	2,625	2,625
Number of person-year observations	22,766	22,766	22,766	22,766	22,766	22,766

*** p<0.01, ** p<0.05, * p<0.10

Note: All models include a constant and controls for age, age squared, age cubed, year of birth, gender, and columns (3)-(6) include controls for birth order, region of birth, and indices intended to capture long-term planning horizon (coefficients suppressed to conserve space).

Table 4. Race & SES Differences in Mid to Late-life Health (Over 55):Importance of Young Adult Neighborhood & Family Background

(Dependent variable: general health status in adulthood)

4-Level Hierarchical Random Effects Interval Regression Model: 100pt-scale, 100=perfect healtl

	Uncond'l model	Raw race gap	Controls for Childhood SES	Controls for Childhood SES + Young Adult Family	Controls for Childhood SES + Young Adult Nhood + Fam	Controls for Childhood SES + Young Adult backgrd + Contemporaneous NHood
Childhood factors	(1)	(2)	(3)	(4)	(5)	(6)
Black		-15.2231***	-9.8758***	-5.2408***	-3.6323***	-3.8721***
Non-Hispanic white (reference category)		(0.3902)	(0.3946)	(0.3851)	(0.4049)	(0.4058)
Childhood poverty			-3.7472***	-3.0631***	-2.9577***	-2.9434***
Non-poor (reference category)			(0.2280)	(0.2197)	(0.2193)	(0.2196)
Mother's education:						
High school dropout			-4.3180***	-2.3982***	-2.4046***	-2.4144***
High school graduate (reference category)			(0.2284)	(0.2227)	(0.2241)	(0.2244)
College educated			2.3119***	0.9967***	1.2704***	1.2114***
			(0.3532)	(0.3407)	(0.3399)	(0.3404)
Father's education:						
High school dropout			-2.5724***	-1.6665***	-1.7798***	-1.7860***
High school graduate (reference category)			(0.2508)	(0.2414)	(0.2405)	(0.2409)
College educated			1.4684***	0.3004	0.1198	0.1359
			(0.3708)	(0.3593)	(0.3574)	(0.3578)
Father's occupation:						
White collar			1.4200***	0.4631+	0.7858**	0.7271**
Blue collar (reference category)			(0.3445)	(0.3334)	(0.3324)	(0.3329)
Grew up on farm			-0.7676***	0.8425***	0.7363***	0.6976**
Grew up in large, urban MSA (ref. category)			(0.2894)	(0.2743)	(0.2764)	(0.2768)
Grew up in small town			-2.4768***	-1.6318***	-1.7885***	-1.7948***
			(0.2563)	(0.2407)	(0.2419)	(0.2422)
Grew up in different places			0.0192	-0.8392	0.1602	0.1722
			(0.9629)	(0.9055)	(0.8995)	(0.9004)
Young Adulthood factors						
Family income-to needs ratio (avg during 1967-1972), spline:						
Income-to-needs ratio*ratio is <1				14.4338***	15.1875***	15.4341***
				(1.5567)	(1.5577)	(1.5593)
Income-to-needs ratio* ratio is 1 to 3				2.6004***	2.0381***	2.0775***
				(0.1937)	(0.1938)	(0.1940)
Income-to-needs ratio* ratio is >3				0.2545***	0.2012**	0.2175***
				(0.0844)	(0.0837)	(0.0838)
Educational attainment:						
High school dropout				-8.0484***	-7.2543***	-7.2735***
High school graduate (reference category)				(0.2871)	(0.2881)	(0.2885)
Some college				0.9893***	1.1771***	1.1769***
				(0.2626)	(0.2628)	(0.2632)
College graduate or higher				5.0618***	5.4568***	5.4627***
				(0.2717)	(0.2715)	(0.2718)
No Private HI coverage, 1968-1972				-3.2845***	-2.5922***	-2.6260***
				(0.2183)	(0.2192)	(0.2194)
Smoked cigarettes at some point, 1968-1972				-2.7553***	-2.6894***	-2.7168***
				(0.2195)	(0.2180)	(0.2182)
Annual alcohol expenditures (in $100's), 5-year average 1968-1972				0.0341**	0.0198	0.0197
				(0.0155)	(0.0155)	(0.0155)

Table 4 (cont'd). Race & SES Differences in Mid to Late-life Health (Over 55): Importance of Young Adult Neighborhood & Family Background

(Dependent variable: general health status in adulthood)
4-Level Hierarchical Random Effects Interval Regression Model: 100pt-scale, 100=perfect healtl

	Uncond'l model	Raw race gap	Controls for Childhood SES	Controls for Childhood SES + Young Adult Family	Controls for Childhood SES + Young Adult Nhood + Fam	Controls for Childhood SES + Young Adult backgrd + Contemporaneous NHood
Young Adult Neighborhood factors	(1)	(2)	(3)	(4)	(5)	(6)
Neighborhood poverty rate (1970), spline:						
Low poverty neighborhood (reference category)						
Medium poverty neighborhood					-2.7542***	-3.0075***
					(0.3558)	(0.3568)
(Neighborhood poverty rate - 20)* rate is 10 to 30%					-1.4913***	-1.7541***
					(0.4985)	(0.4996)
High poverty neighborhood					-6.3390***	-6.4341***
					(0.5921)	(0.5929)
Neighborhood crime problem					-1.2592***	-1.2929***
					(0.3713)	(0.3717)
Residential segregation dissimilarity index, 1970 (MSA)					-4.9635***	-5.2839***
					(1.8422)	(1.8447)
Neighborhood housing quality index					-0.7885***	-0.7852***
					(0.0724)	(0.0725)
Neighborhood connectedness to informal sources of help					0.7270***	0.7260***
					(0.0750)	(0.0751)
Average Aspirations index in neighborhood					0.5967***	0.5995***
					(0.0677)	(0.0678)
Contemporaneous Neighborhood						
Neighborhood poverty rate						0.5512***
						(0.0459)
Random Effects, Unmeasured (Std Dev)						
Young Adult Neighborhood component	10.5244***	9.0692***	6.9812***	5.2475***	4.8412***	4.8621***
	(0.1718)	(0.1865)	(0.2142)	(0.2416)	(0.2576)	(0.2575)
Family component	5.0742***	5.5954***	5.5354***	4.8759***	4.5505***	4.5092***
	(0.3605)	(0.3277)	(0.3300)	(0.3473)	(0.3707)	(0.3749)
Individual component	15.0163***	15.0116***	15.0581***	14.8695***	14.9012***	14.9303***
	(0.1070)	(0.1062)	(0.1065)	(0.1040)	(0.1043)	(0.1045)
Transitory error component	7.7180***	7.7180***	7.7186***	7.7175***	7.7177***	7.7130***
	(0.0122)	(0.0122)	(0.0122)	(0.0122)	(0.01218)	(0.0122)
Log-likelihood	-875160.66	-874415.73	-873167.25	-871477.14	-871153.91	-871078.29
Number of counties	306	306	306	306	306	306
Number of neighborhoods	1,129	1,129	1,129	1,129	1,129	1,129
Number of families	1,460	1,460	1,460	1,460	1,460	1,460
Number of individuals	2,019	2,019	2,019	2,019	2,019	2,019
Number of person-year observations	13,355	13,355	13,355	13,355	13,355	13,355

*** p<0.01, ** p<0.05, * p<0.10

Note: All models include a constant and controls for age, age squared, age cubed, year of birth, gender, and columns (3)-(6) include controls for birth order, region of birth, and indices intended to capture long-term planning horizon (coefficients suppressed to conserve space).

Table 5. Spousal and Young Adult Neighbor Correlations in Later-Life Health Status

	All Adulthood yrs (35+)		Age 35-55		Over 55	
	Spousal Correlation	**Young Adult Neighbor Correlation**	**Spousal Correlation**	**Young Adult Neighbor Correlation**	**Spousal Correlation**	**Young Adult Neighbor Correlation**
Unconditional	0.4598	0.3313	0.4718	0.3485	0.3771	0.3060
	(0.0067)	(0.0077)	(0.0068)	(0.0085)	(0.0089)	(0.0089)
Adjusted* (net of residential sorting of HHs w/similar family bckgrd)	--	0.2677				
Conditional, control for childhood SES	0.3605	0.1988	0.3739	0.2116	0.2593	0.1592
	(0.0075)	(0.0083)	(0.0076)	(0.0098)	(0.0098)	(0.0095)
Conditional, control for childhood + young adult family/neighborhood factors	0.2760	0.1287	0.3003	0.1409	0.1658	0.0880
	(0.0080)	(0.0079)	(0.0081)	(0.0094)	(0.0104)	(0.0093)

*To compute the adjusted neighbor correlations, we first estimated within-neighborhood estimates of the effects of family income, education, race, family structure, health insurance coverage, health behaviors, connectedness to informal sources of help and housing quality on health in adulthood. Then, we used the within-neighborhood estimates of the later-life health effects of this array of family characteristics to assess how much of the raw neighbor correlation is due to young adult neighbors having similar (observable) family characteristics as opposed to neighborhood effects *per se*. The estimation procedures are described in detail in the methods section of the paper. (The full model results used to compute the adjusted neighbor correlations are not shown to conserve space, but are available from the authors upon request).

Table 6. Estimated Effect of Living in High Poverty Neighborhood during Young Adulthood on Later-life Health for a Proportional Correlation Model with Varying Values of the Relative Correlation

Relative Correlation	Estimated Effect of High Poverty Neighborhood during Young Adulthood (reference cat: Low Poverty Neighborhood)
0 (exogeneity)	-4.7344***
	(0.1198)
0.2	-5.1847***
	(0.2804)
0.4	-2.4631***
	(0.2194)
0.8	-13.1834***
	(0.2136)
1	-18.6636***
	(0.7760)

Data Appendix Table A0.

Measures	Data Source	Year(s) collected	Survey Question	Definition
General Health Status	PSID	Adulthood:1984-2005	“Would you say your health in general is excellent, very good, good, fair, or poor?”	--
Childhood SES	PSID	Childhood circumstances retrospectively collected in 1968-1972 waves	“Were your parents poor when you were growing up, well off, or avg?""How much education did your parents have?""Did you grow up on a farm, in a large city, small town?”	--
Neighborhood Poverty Rate	1970-2000 Census	Young adult neighborhood: 1970 Census; Mid-to-Late Life neighborhood: 1980-2000 (linearly interpolate for non-census years)	PSID respondent's residential location (1968-2005) matched to decennial census tract info	low poverty neighborhood (<10% poor); medium poverty neighborhood (10-30%); high poverty neighborhood (>30%)
Racial Residential Segregation	1970 Census	Young adulthood: 1970 Census	Black-white dissimilarity index$_{MSA}$: b_{it} & w_{it} = # of black & white individuals in neighborhood i at time t; B_t & W_t = total # black & white individuals in MSA.	$\frac{1}{2} * \sum_{i=1}^{n} \left\| \frac{b_{it}}{B_t} - \frac{w_{it}}{W_t} \right\|$
Neighborhood/Housing Quality	PSID	Young adulthood:1975	Self-reports: whether there exist housing structural, plumbing or insulation problems, or burglary, robbery, assault, drug use problems, or too few police in neighborhood in which they live.	High crime neighborhood=avg response among all PSID households who live in same neighborhood report major crime-related problems; housing insulation/plumbing problems=avg response among all PSID households who live in same neighborhood
Neighborhood Connectedness to informal sources of support	PSID	1968-1972	Index (0-9) of Connectedness to Potential Sources of Help (constructed from survey responses): Attends church once a month or more; # of neighbors known by name; Has relatives within walking distance; Goes to organizations once a month or more.	Neighborhood-level measures obtained by computing avg index score based on responses among all PSID HHs who live in same neighborhood.
Neighborhood Aspirations	PSID	1968-1972	Index (0-9) (constructed from survey responses): "Plans to move(purposive move);Has high educational aspirations for children; Says it is important to make own decisions on a job; Is willing to move for even a moderately better job; Has plans to try for a new job; Doesn't like a job where told what to do; Wanted to work more hours than did or Did not want to work fewer hours than did; Expects things to happen for better; Likes to do things difficult or challenging; Would rather have a job with good chance for making more, even if don't like it; Spends time figuring out ways to get more money"	Neighborhood-level measures obtained by computing avg reponse among all PSID HHs who live in same neighborhood.

Table A1. Descriptive Statistics by Race

	All (N=2,730)	Black (N=887)	White (N=1,782)
Adult Health Status:			
Excellent	0.17	0.08	0.21
Very Good	0.29	0.18	0.34
Good	0.31	0.33	0.30
Fair	0.16	0.29	0.11
Poor	0.07	0.13	0.04
Age (range: 37-77)	53.0	53.5	52.9
Year born (range: 1928-1948)	1939	1939	1939
Female	0.54	0.65	0.53
Childhood factors			
Childhood poverty	0.38	0.72	0.35
Mother's education:			
High school dropout	0.49	0.78	0.46
High school graduate	0.39	0.18	0.41
College educated	0.11	0.05	0.12
Father's education:			
High school dropout	0.64	0.83	0.62
High school graduate	0.24	0.14	0.25
College educated	0.12	0.03	0.13
Father's occupation:			
White collar	0.11	0.04	0.11
Grew up on farm	0.31	0.38	0.30
Grew up in large, urban metropolitan area	0.33	0.31	0.33
Grew up in small town	0.35	0.30	0.36
Young Adulthood factors			
Income-to-needs ratio (5-yr avg, 1968-1972):			
<1 (poverty)	0.05	0.24	0.03
1-3	0.51	0.53	0.50
>3	0.44	0.23	0.47
Educational attainment:			
High school dropout	0.19	0.37	0.16
High school graduate	0.38	0.37	0.38
Some college	0.20	0.18	0.20
College graduate or higher	0.24	0.08	0.25
No private health insurance, 1968-1972	0.38	0.58	0.36
Health behaviors (1997 $):			
Smoked cigarettes at some point, 1968-1972	0.71	0.78	0.70
Alcohol consumption (5-yr avg, 1968-1972)	$416	$461	$413
Young adult neighborhood variables:			
Neighborhood poverty:			
High poverty neighborhood (>30%)	0.04	0.19	0.02
Medium poverty neighborhood (10-30%)	0.17	0.50	0.14
Low poverty neighborhood (<10%)	0.79	0.31	0.84
High crime neighborhood	0.08	0.15	0.07
Residential segregation dissimilarity index$_{MSA}$	0.80	0.78	0.80
N'hood housing quality index	1.00	1.78	0.91
N'hood connectedness to informal sources of help	5.97	6.10	5.98
Neighborhood aspirations index	4.82	5.06	4.80

Note: All descriptive statistics are sample weighted to produce nationally-representative estimates of means. Black-white differences in all family and neighborhood factors are statistically significant.